# 现代应用数学手册

《现代应用数学手册》编委会

## 运筹学与最优化理论卷



清華大學出版社
http://www.tup.tsinghua.edu.cn

《现代应用数学手册》编委会

# 现代应用数学手册

## 运筹学与最优化理论卷

清华大学出版社

（京）新登字 158 号

## 内容简介

本书是运筹学与最优化理论的常备工具书，内容新颖，实用性强，查阅方便。主要介绍运筹学与最优化理论在科学研究、工程技术、经济管理中各种实际问题的数学模型及计算方法。本卷分为 6 章，依次为线性规划、非线性规划、整数规划、动态规划、多目标规划及对策论，并配有丰富的例题，便于读者加深理解、掌握与运用。为便于读者查阅，还附有中文—外文名词索引、外文—中文名词索引。本书可供广大科研人员、工程技术人员、管理人员、大中专院校师生和研究生使用。

**图书在版编目（CIP）数据**

现代应用数学手册：运筹学与最优化理论卷/《现代应用数学手册》编委会编．—北京：清华大学出版社，1997

ISBN 7-302-02760-9

Ⅰ．现… Ⅱ．现… Ⅲ．①应用数学-手册②运筹学③最优化-数学理论

Ⅳ．029-62

中国版本图书馆 CIP 数据核字（97）第 25853 号

出版者：清华大学出版社（北京清华大学校内，邮编 100084）
因特网地址：www.tup.tsinghua.edu.cn

印刷者：北京市清华园胶印厂

发行者：新华书店总店北京科技发行所

开　本：850×1168 1/32　印张：16　字数：413 千字

版　次：1998 年 1 月第 1 版　1998 年 4 月第 1 次印刷

书　号：ISBN 7-302-02760-9/O・190

印　数：0001～4000

定　价：28.00 元

# 《现代应用数学手册》
# 编辑委员会

# 运筹学与最优化理论卷

**责任编委　郑乐宁**

| 章次 | 编者 | 校者 |
| --- | --- | --- |
| 1 | 王燕来 | 陈宝林 |
| 2 | 陈宝林 | 王燕来 |
| 3 | 马仲蕃 | 刘振宏 |
| 4 | 姜启源 | 郑乐宁 |
| 5 | 郑乐宁 | 谭泽光 |
| 6 | 王建华 | 马振华 |

# 序

随着计算机科学技术的飞速发展，人类正进入信息时代.

信息时代是应用数学大发展的时代，人类长期积累起来的知识体系，正面临着第3次数学化. 数学思想，数学方法与数学模型随着计算机的广泛应用，日益渗透到各种行业中去.

当代，除了古典的数学理论（初等数学，微积分学，微分方程，复变函数等）早已得到广泛的应用外，一些比较抽象的现代数学理论（集合论、数理逻辑、范畴论、抽象代数、泛代数、代数几何、拓扑学、泛函分析等）以及一些新兴的数学理论（随机过程、时间序列、运筹学、最优化理论、有限元方法、模糊数学、混沌与分形等）也逐渐地成为社会生产，科学实验，工程技术及经济管理中不可缺少的工具，应用数学的适用范围正在迅速地扩大.

为了满足日益增长的社会需求，清华大学应用数学系《现代应用数学手册》编委会，组织编写了这套多卷集的手册.

本书读者是理、工、医、农、经管等各个领域中的广大工程技术人员、科研人员，大、中专院校的教师、学生、研究生及其他使用数学工具的实际工作者. 其中有些内容对于中学生也是适用的.

编者力求使本书成为一套高质量的工具书，它有下列特点：

(1) **内容“新颖”** 本书力求做到内容现代化，除用现代观点介绍古典内容外，对已出现的新理论、新方法尽量优先选入.

(2) **突出“应用”** 本书在选材上突出数学理论的应用，以通俗易懂的方式着重介绍在现代科学技术等实际领域中应用广泛的数学理论和方法.

（3）**紧密“结合”计算机应用** 为了更有效地应用数学方法解决各种实际问题，广大科技人员迫切要求数学方法与计算机应用相结合，提高工作效率．为此，本书在结合计算机应用方面，给予特别的重视．

（4）**版面设计“合理”，便于迅速查阅** 为方便读者使用，本书采用了一套较为完善的索引体系．除正文中章、节的编号沿用国际通行的十进制编号外，对于重要的定义、定理、例题、公式、图、表等均有编号．读者可以从（1）目录，（2）中文—外文索引，（3）外文—中文索引等三种途径，迅速找到所需资料．此外，本书对载入的外国科学家人名，尽量采用“名从主人”的原则．

（5）**数学符号力求“统一”与国际化** 鉴于目前国内各种文献、书籍中使用的数学符号不够统一与国际化，增加了读者阅读时的困难．本书除按国家标准 GB3102-93 外，兼用国际数学界权威著作《数学大百科辞典》(Encyclopedic Dictionary of Mathematics，EDM)中的符号为标准．对于不在上述文献中的其他新符号，则选用较为流行者．

本手册各卷内容独立完整，便于个人读者与团体读者按需选购．当前应用数学急剧发展，编委会在条件成熟的时候，还将增出新卷．

本书的编撰是与清华大学应用数学系领导，特别是萧树铁教授的热心支持，编辑委员会各位编委的通力协作，校内外的许多教师、科研工作者的大力支持分不开的，编者深致谢意．

在编辑出版过程中，还得到清华大学出版社的热情支持．

本书从编撰到出版，历尽艰辛．饮水思源，编者还要感谢本书的发起人，清华大学应用数学系陆璇教授，北京出版社李利军编辑及已故的北京出版社社长王政人先生．

最后，编者还要对夫人王华敏表示谢忱，没有她的深刻理解、热情支持与持久的帮助，本书也难以问世.

主编　马振华

1997年于清华园

# 数学符号表

| | |
|---|---|
| $\forall$ | 全称量词 |
| $\exists$ | 存在量词 |
| $\Leftrightarrow$ | 等价 |
| $\lfloor\ \rfloor$/[ ] | 底($\lfloor\alpha\rfloor$是不超过实数 $\alpha$ 的最大整数) |
| $\lceil\ \rceil$ | 顶($\lceil\alpha\rceil$是不小于实数 $\alpha$ 的最小整数) |
| $O$ | Landau 符号(当 $x\to\alpha$ 时 $\lvert f(x)/g(x)\rvert$ 有界,<br>记作 $f(x)=O(g(x))$) |
| $o$ | Landau 符号(当 $x\to\alpha$ 时 $\lvert f(x)/g(x)\rvert\to 0$,<br>记作 $f(x)=o(g(x))$) |
| $\sim$ | 等价(当 $x\to\alpha$ 时 $f(x)/g(x)\to 1$) |
| $:=$ | 定义与赋值号($k:=k+1$) |
| $>^l$ | 字典序($\boldsymbol{x}>^l\boldsymbol{y}$,$x$ 按字典序大于 $y$/$x$ 优超于 $y$) |
| $\rho(A)$ | 矩阵 $A$ 的谱半径($\rho(A)=\max\limits_{1\leqslant i\leqslant n}\lvert\lambda_i\rvert$) |
| $\nabla f(\boldsymbol{x})$ | $f(\boldsymbol{x})$的梯度 |
| $\nabla^2 f(\boldsymbol{x})$ | $f(\boldsymbol{x})$的 Hesse 矩阵 |
| $s\parallel z^{(i)}$ | 在 $n$ 维向量$(x^{(1)},\cdots,x^{(i)},\cdots,x^{(n)})$中,用 $z^{(i)}$代替 $x^{(i)}$:<br>$s\parallel z^{(i)}\triangleq(x^{(1)},\cdots,x^{(i-1)},z^{(i)},x^{(i+1)},\cdots,x^{(n)})$) |
| lex min | 按字典序最小 |

# 目　录

# 1 线性规划

## 1.1 线性规划问题

### 1.1.1 引言

**线性规划**(linear programming)是运筹学中产生较早、应用广泛的一个分支.早在20世纪30年代,Канторович研究并发表了《生产组织与计划的数学方法》,其中论述的就是线性规划问题. 1947年G. B. Dantzig提出了单纯形法.其后在计算机上的成功实现使得应用线性规划解决的问题迅速增加.线性规划已广泛用于国防、科技、经济、工业、农业、环境工程、教育及社会科学中.

线性规划是研究在一组线性约束之下,某个线性函数的最小值或最大值问题.

**例 1.1.1** 某工厂有三种原料 $B_1$,$B_2$ 和 $B_3$,其储量分别为 170 kg、100 kg 和 150 kg.现用来生产两种产品 $A_1$ 和 $A_2$.已知每生产 1 kg $A_1$ 产品需要原料 $B_1$ 5 kg,原料 $B_2$ 2 kg,原料 $B_3$ 1 kg.每生产 1 kg$A_2$ 产品需要原料 $B_1$ 2 kg,原料 $B_2$ 3 kg,原料 $B_3$ 5 kg.又知 $A_1$ 产品每 kg 利润为 10 元,$A_2$ 产品每 kg 利润为 18 元.问在工厂现有资源条件下,应如何安排生产,才使工厂获得最大利润?

**解** 设安排 $A_1$,$A_2$ 产品的产量分别为 $x_1$,$x_2$ kg.则该问题的数学模型为

$$
\begin{aligned}
\max \quad & 10x_1 + 18x_2; \\
\text{s.t.} \quad & 5x_1 + 2x_2 \leqslant 170, \\
& 2x_1 + 3x_2 \leqslant 100, \\
& x_1 + 5x_2 \leqslant 150, \\
& x_1 \geqslant 0, \\
& x_2 \geqslant 0.
\end{aligned}
$$

这是一个线性规划问题(linear programming problem)，简记为(LP).

一般地，线性规划问题的数学模型为

$$
\left.
\begin{aligned}
\min/\max \quad & f(x_1,\cdots,x_n)=c_1x_1+\cdots+c_nx_n; \\
\text{s.t.} \quad & a_{11}x_1+\cdots+a_{1n}x_n\leqslant b_1 \quad (\text{或}\geqslant b_1;\text{或}=b_1), \\
& \cdots\cdots \qquad\qquad\qquad\qquad\qquad\qquad , \\
& a_{m1}x_1+\cdots+a_{mn}x_n\leqslant b_m \quad (\text{或}\geqslant b_m;\text{或}=b_m), \\
& x_j\geqslant 0 \quad (j=1,\cdots,n).
\end{aligned}
\right\} \quad (1.1)
$$

其中，$x_1,\cdots,x_n$ 称为**决策变量**(decision variables)，是线性规划问题中要求解的变量. $f(x_1,\cdots,x_n)=c_1x_1+\cdots+c_nx_n$ 称为**目标函数**(objective function). 常数 $c_1,\cdots,c_n$，称为**费用系数**(cost coefficients). $a_{i1}x_1+\cdots+a_{in}x_n\leqslant b_i$(或$\geqslant b_i$;或$=b_i$) $(i=1,\cdots,m)$称为第 $i$ 个**约束**(constraint). 系数 $a_{ij}$组成的矩阵

$$
\boldsymbol{A}=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}
$$

称为**约束矩阵**(constraint matrix). 列向量 $\boldsymbol{b}=(b_1,\cdots,b_m)^{\mathrm{T}}$ 称为**右端向量**(right-hand-side vector). 条件 $x_j\geqslant 0(j=1,\cdots,n)$称为**非负约束**. 本书中约束条件记为 s.t. 是 subject to 的缩写，意思是**满足约束条件**.

### 1.1.2 线性规划问题的标准形式

线性规划问题的标准形式表示为：

$$\left.\begin{aligned}
\min\quad & c_1x_1 + c_2x_2 + \cdots + c_nx_n ; \\
\text{s.t.}\quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\
& \quad\vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \quad\ \vdots , \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \\
& x_1 \geqslant 0,\ x_2 \geqslant 0,\ \cdots\ x_n \geqslant 0.
\end{aligned}\right\} \tag{1.2}$$

其中，$x_i$ 为决策变量，$c_i$，$a_{ij}$和 $b_i$ 为实常数. 并且，$b_i \geqslant 0$.

利用向量和矩阵符号可以简写为：

$$\left.\begin{aligned}
\min\quad & \boldsymbol{cx}; \\
\text{s.t.}\quad & \boldsymbol{Ax} = \boldsymbol{b}, \\
& \boldsymbol{x} \geqslant \boldsymbol{0}.
\end{aligned}\right\} \tag{1.3}$$

其中，$\boldsymbol{A}$ 为矩阵，$\boldsymbol{b}$，$\boldsymbol{c}$，$\boldsymbol{x}$ 均为向量：

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

$$\boldsymbol{c} = (c_1, c_2, \cdots, c_n),$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^{\mathrm{T}},$$

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}.$$

在各种实际问题中，列出的模型不尽相同，目标函数可能是求最小值或最大值；约束条件可能是等式或不等式；变量可能有限制，或没有限制. 但是，任何一种线性规划模型都可以等价地转换为标准形式.

(1) 目标函数的转换　若原问题是求 max $\boldsymbol{cx}$，则可以转换成

求 min $(-cx)$.

(2) 约束条件的转换　若某约束条件是不等式

$$\sum_{j=1}^{n} a_{ij}x_j \leqslant b_i \quad \left( \sum_{j=1}^{n} a_{ij}x_j \geqslant b_i \right),$$

则可以转换成为

$$\begin{cases} \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \left( \sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i \right), \\ x_{n+i} \geqslant 0. \end{cases}$$

其中,引进的新变量 $x_{n+i}$ 称为**松弛变量**(slack variable). 在目标函数中与松弛变量相对应的费用系数取 0 值.

若等式约束 $\sum_{j=1}^{n} a_{ij}x_j = b_i$ 中的 $b_i < 0$,则可将该等式的两端同乘以$(-1)$转换为 $-\sum_{j=1}^{n} a_{ij}x_j = -b_i$,此时,$(-b_i) > 0$.

(3) 变量的非负约束　若某个变量的约束为 $x_j \leqslant \beta_j$(或 $x_j \geqslant \beta_j$),则可令 $x_j' = \beta_j - x_j$(或 $x_j' = x_j - \beta_j$),于是便有 $x_j' \geqslant 0$;若某个变量 $x_j$ 没有限制,则可令 $x_j = x_j' - x_j''$,并增加约束 $x'_j \geqslant 0, x_j'' \geqslant 0$.

**例 1.1.2**　将下列线性规划模型转化为标准形式

$$\begin{aligned} \min \quad & -x_1 + 2x_2 - 3x_3; \\ \text{s.t.} \quad & x_1 + x_2 + x_3 \leqslant 7, \\ & x_1 - x_2 + x_3 \geqslant 2, \\ & 3x_1 - x_2 - 2x_3 = -5, \\ & x_1 \geqslant 0, \quad x_2 \geqslant 0. \end{aligned}$$

**解**　对前两个约束分别引进松弛变量 $x_6 \geqslant 0, x_7 \geqslant 0$,且规定其相应的费用系数 $c_6 = c_7 = 0$. 对等式约束的两端同乘以$(-1)$. 再令 $x_3 = x_4 - x_5$,并增加约束 $x_4 \geqslant 0, x_5 \geqslant 0$. 这样,给定的线性规划模型转化为标准形式:

$$
\begin{aligned}
\min \quad & -x_1+2x_2-3x_4+3x_5 \; ; \\
\text{s.t.} \quad & x_1+x_2+x_4-x_5+x_6=7, \\
& x_1-x_2+x_4-x_5-x_7=2, \\
& -3x_1+x_2+2x_4-2x_5=5, \\
& x_1\geqslant 0,\ x_2\geqslant 0, \\
& x_4\geqslant 0,\ x_5\geqslant 0, \\
& x_6\geqslant 0,\ x_7\geqslant 0.
\end{aligned}
$$

下面给出线性规划问题解的概念.

**定义 1.1.3** (1.15)中,满足约束条件的向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 称为(1.15)的**可行解**(feasible solution)或**可行点**(feasible point).可行点的集合,称为**可行域**(feasible region).记作

$$S=\{\boldsymbol{x}\in E^n \mid \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geqslant \boldsymbol{0}\}.$$

使目标函数取得最小值的可行解称为**最优解**(optimal solution).

### 1.1.3 线性规划的图解法

如果线性规划问题只含有两个决策变量,那么,它的可行域可在平面上画出,因而可用图解法求最优解.举例说明.

**例 1.1.4** 用图解法解下列线性规划

$$
\begin{aligned}
\min \quad & -x_1-4x_2 \; ; \\
\text{s.t.} \quad & x_1+x_2\leqslant 4, \\
& -x_1+x_2\leqslant 2, \\
& x_1\geqslant 0,\quad x_2\geqslant 0.
\end{aligned}
$$

**解** 这个问题的可行域 $S$ 如图 1.1 所示.目标函数等值线的方程为 $-x_1-4x_2=z$.因为要找的最优解在可行域内使目标函数具有最小值,所以,让等值线 $-x_1-4x_2=z$ 沿着它的负法线方向 $-\boldsymbol{n}=(1,4)^{\mathrm{T}}$ 在可行域内尽量地平行移动.即,移动到图 1.1 中点 $(1,3)^{\mathrm{T}}$ 的位置.如果再移动就与可行域不相交了,于是点 $(1,3)^{\mathrm{T}}$

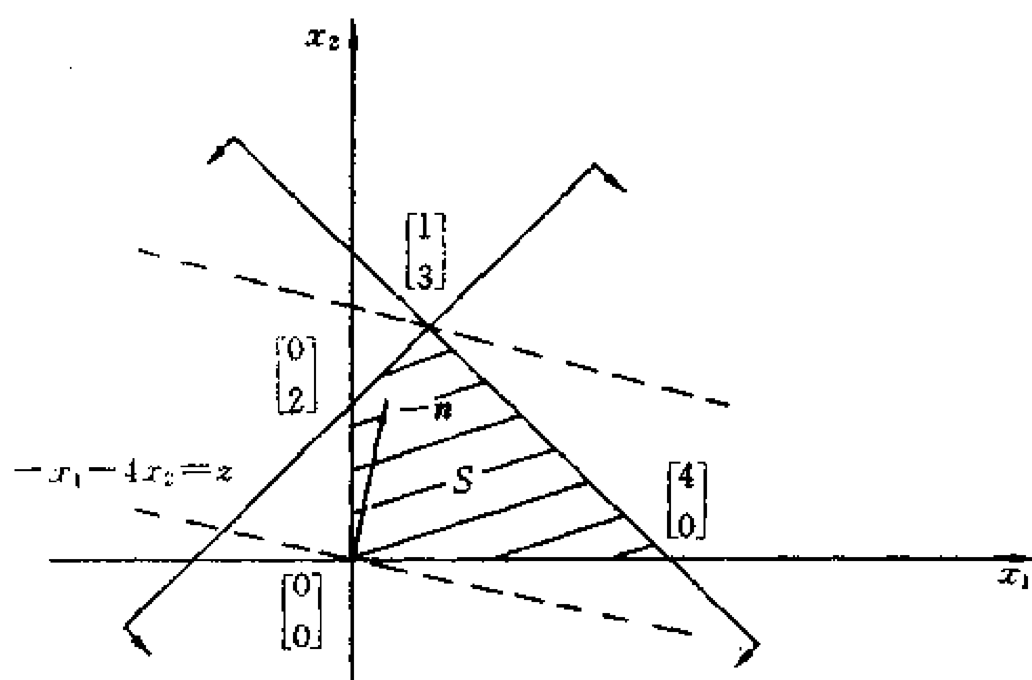



图 1.1

就是最优解，目标函数最优值为$-13$.

### 1.1.4 线性规划的性质

首先给出线性规划问题基本可行解的概念.

**定义 1.1.5** 考虑(1.3)，设 $\boldsymbol{B}$ 是矩阵 $\boldsymbol{A}$ 中的一个 $m$ 阶满秩子方阵，则称 $\boldsymbol{B}$ 为一个**基**(basis). $\boldsymbol{A}$ 中剩余元素组成的子阵记为 $\boldsymbol{N}$，即 $\boldsymbol{A}=(\boldsymbol{B},\boldsymbol{N})$. 再把 $\boldsymbol{x}$ 的分量相应地分作两部分，记成 $\boldsymbol{x}_B$ 和 $\boldsymbol{x}_N$，$\boldsymbol{x}_B$ 的分量与 $\boldsymbol{B}$ 的列对应，称为**基变量**(basic variables). $\boldsymbol{x}_N$ 的分量与 $\boldsymbol{N}$ 中的列对应，称为**非基变量**(nonbasic variables). 在约束 $\boldsymbol{Ax}=\boldsymbol{b}$ 中令所有的非基变量取值为零时，得到的解 $\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\\boldsymbol{x}_N\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}^{-1}\boldsymbol{b}\\\boldsymbol{0}\end{bmatrix}$，称为相应于 $\boldsymbol{B}$ 的**基本解**(basic solution).

**定义 1.1.6** 基本解的基变量都取非负值时，即满足 $\boldsymbol{x}_B\geqslant\boldsymbol{0}$

的基本解称为**基本可行解**(basic feasible solution). 相应的基 $\boldsymbol{B}$ 称为**可行基**(feasible basis).

**注意**,有 $m$ 个约束方程和 $n$ 个变量的标准形式线性规划问题必有有限个基本解,其最大个数为

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

基本可行解的最大组数也小于或等于此式的值.

**定义 1.1.7** 若基本可行解的所有基变量都取正值,则称它是**非退化的**(non-degenerate);若有取零值的基变量,则称它是**退化的**(degenerate). 一个线性规划问题,若它的所有基本可行解都是非退化的,则称该问题是**非退化**的.

**例 1.1.8** 选择不同的基,求下列线性规划的基本解.

$$\begin{aligned}
\min \quad & -x_1 - 4x_2 ; \\
\text{s.t.} \quad & x_1 + x_2 + x_3 = 4, \\
& -x_1 + x_2 + x_4 = 2, \\
& x_1, x_2, x_3, x_4 \geqslant 0.
\end{aligned}$$

**解** 问题的约束矩阵为

$$\boldsymbol{A} = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \boldsymbol{a}_3\ \boldsymbol{a}_4] = \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix}.$$

下面选择不同的基,并分别求出所对应的基本解:

$$\boldsymbol{B}_1 = [\boldsymbol{a}_1\ \boldsymbol{a}_2] = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix},\text{求得基本解为}$$

$$\boldsymbol{x}_1 = (x_1, x_2, x_3, x_4)^{\mathrm{T}} = (1, 3, 0, 0)^{\mathrm{T}}.$$

$$\boldsymbol{B}_2 = [\boldsymbol{a}_1\ \boldsymbol{a}_3] = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix},\text{求得基本解为}$$

$$\boldsymbol{x}_2 = (x_1, x_2, x_3, x_4)^{\mathrm{T}} = (-2, 0, 6, 0)^{\mathrm{T}}.$$

$$\boldsymbol{B}_3 = [\boldsymbol{a}_1\ \boldsymbol{a}_4] = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix},\text{求得基本解为}$$

$$\boldsymbol{x}_3=(x_1,x_2,x_3,x_4)^{\mathrm{T}}=(4,0,0,6)^{\mathrm{T}}.$$

$\boldsymbol{B}_4=[\boldsymbol{a}_2\ \boldsymbol{a}_3]=\begin{bmatrix}1&1\\1&0\end{bmatrix}$,求得基本解为

$$\boldsymbol{x}_4=(x_1,x_2,x_3,x_4)^{\mathrm{T}}=(0,2,2,0)^{\mathrm{T}}.$$

$\boldsymbol{B}_5=[\boldsymbol{a}_2\ \boldsymbol{a}_4]=\begin{bmatrix}1&0\\1&1\end{bmatrix}$,求得基本解为

$$\boldsymbol{x}_5=(x_1,x_2,x_3,x_4)^{\mathrm{T}}=(0,4,0,-2)^{\mathrm{T}}.$$

$\boldsymbol{B}_6=[\boldsymbol{a}_3\ \boldsymbol{a}_4]=\begin{bmatrix}1&0\\0&1\end{bmatrix}$,求得基本解为

$$\boldsymbol{x}_6=(x_1,x_2,x_3,x_4)^{\mathrm{T}}=(0,0,4,2)^{\mathrm{T}}.$$

**注意**,与基 $\boldsymbol{B}_1,\boldsymbol{B}_3,\boldsymbol{B}_4,\boldsymbol{B}_6$ 对应的基本解 $\boldsymbol{x}_1,\boldsymbol{x}_3,\boldsymbol{x}_4,\boldsymbol{x}_6$ 是基本可行解;与基 $\boldsymbol{B}_2,\boldsymbol{B}_5$ 对应的 $\boldsymbol{x}_2,\boldsymbol{x}_5$ 是基本解,但不是基本可行解,因为 $\boldsymbol{x}_2$ 的第 1 个分量是负数,$\boldsymbol{x}_5$ 的第 4 个分量是负数.此外,由于所有的基本可行解都是非退化的,故该问题是非退化的.

下面给出标准形式线性规划问题的几个重要性质.

**定理 1.1.9** 线性规划问题的可行域为凸集.

**定理 1.1.10** $\boldsymbol{x}$ 是式(1.3)的基本可行解的充要条件是它是可行域的极点.

**定理 1.1.11** 若式(1.3)有可行解,则必有基本可行解.

**定理 1.1.12** 若式(1.3)有有限最优值,则必在可行域的极点处达到.

**定理 1.1.13** 式(1.3)有有限最优值的充要条件是对可行域的所有极方向 $\boldsymbol{d}_j$,均有 $\boldsymbol{c}\boldsymbol{d}_j\geqslant 0$.

## 1.2 单纯形法

### 1.2.1 单纯形法原理

对于标准形式的线性规划问题

$$\min \quad f(x)=cx; \tag{1.4}$$

$$\text{s.t.} \quad Ax=b, \tag{1.5}$$

$$x \geqslant 0 \tag{1.6}$$

若有有限最优值，则目标函数的最优值必在某一基本可行解处达到．因而只要在基本可行解中寻找最优解．**单纯形法**(simplex method)的基本思想就是先找一个基本可行解．检验是否为最优解．否则，再找一个使目标函数值有改进的基本可行解，进行检验．反复进行迭代，直至找到最优解，或判定问题无界(即无有限最优值)．

设找到初始基本可行解 $\bar{x}$，可行基为 $B$，非基矩阵为 $N$，即 $A=(B\ N)$．于是，$\bar{x}=\begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}=\begin{bmatrix} \bar{b} \\ 0 \end{bmatrix}$．相应地，目标函数值为 $\bar{f}=c\bar{x}=[c_B\ c_N]\begin{bmatrix} \bar{b} \\ 0 \end{bmatrix}=c_B\bar{b}$ 其中，$c_B$ 是 $c$ 中与基变量对应的分量组成的 $m$ 维行向量．

再设任意可行解 $x=\begin{bmatrix} x_B \\ x_N \end{bmatrix}$，由(1.5)得到

$$x_B=B^{-1}b-B^{-1}Nx_N=\bar{b}-B^{-1}Nx_N. \tag{1.7}$$

相应的目标函数值为

$$f=cx=c_B\bar{b}-(c_BB^{-1}N-c_N)x_N. \tag{1.8}$$

若记 $A=[a_1\ a_2 \cdots a_n]$，于是

$$f=\bar{f}-\sum_{j\in N_B}(c_BB^{-1}a_j-c_j)x_j, \tag{1.9}$$

其中，$N_B$ 是非基变量的下标集．记

$$z_j-c_j=c_BB^{-1}a_j-c_j, \tag{1.10}$$

称为**检验数**(criterion)．于是有

$$f=\bar{f}-\sum_{j\in N_B}(z_j-c_j)x_j. \tag{1.11}$$

变换后的问题叙述如下：

$$\min \quad f = \bar{f} - \sum_{j \in N_B} (z_j - c_j) x_j; \tag{1.12}$$

$$\text{s.t.} \quad \boldsymbol{x}_B + \boldsymbol{B}^{-1} \boldsymbol{N} \boldsymbol{x}_N = \bar{\boldsymbol{b}}, \tag{1.13}$$

$$\boldsymbol{x} \geqslant \boldsymbol{0}. \tag{1.14}$$

其中，$\bar{f}$是基本可行解 $\bar{\boldsymbol{x}}$ 所对应的目标函数值.

假设(1.12)～(1.14)对应的基本可行解是非退化的，则有下面定理.

**定理 1.2.1** 若(1.12)中所有的 $z_j - c_j \leqslant 0$，则 $\bar{\boldsymbol{x}}$ 为问题(1.4)～(1.6)的最优解，记为 $\boldsymbol{x}^*$.

**定理 1.2.2** 若(1.12)中有 $z_k - c_k > 0, k \in N_B$，且相应的 $\boldsymbol{B}^{-1}\boldsymbol{a}_k \leqslant 0$，则问题(1.4)～(1.6)有无界的最优值.

**定理 1.2.3** 若(1.12)中有 $z_k - c_k > 0, k \in N_B$，且 $\bar{\boldsymbol{a}}_k = \boldsymbol{B}^{-1}\boldsymbol{a}_k$ 至少有一个正分量，则能找到基本可行解 $\hat{\boldsymbol{x}}$，使目标函数值下降，即 $\boldsymbol{c}\hat{\boldsymbol{x}} < \boldsymbol{c}\bar{\boldsymbol{x}}$.

上述三个定理给出了单纯形法的基本思路：首先找一个基本可行解 $\bar{\boldsymbol{x}}$，对它进行检验. 若与之相应的所有检验数 $z_j - c_j \leqslant 0$，则 $\bar{\boldsymbol{x}}$ 为最优解；若有 $z_k - c_k > 0$，且相应的 $\boldsymbol{B}^{-1}\boldsymbol{a}_k \leqslant 0$，则问题无界；若 $z_k - c_k > 0$，且 $\boldsymbol{B}^{-1}\boldsymbol{a}_k$ 含有正分量，则可以求出一个新的基本可行解 $\hat{\boldsymbol{x}}$，使目标函数值减小. 然后，再检验最优性，再迭代. 由于基本可行解的个数是有限的，故经过有限次迭代后一定能找到最优解或判定问题无界.

为了求出一个新的基本可行解 $\hat{\boldsymbol{x}}$，可取正检验数中最大的 $z_k - c_k$ 的下标 $k$ 所对应的列向量 $\boldsymbol{a}_k$ 进入基，使相应的非基变量 $x_k$ 取正值变为基变量. 令

$$x_k = \min \left\{ \frac{\bar{b}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik} > 0 \right\} \triangleq \frac{\bar{b}_r}{\bar{a}_{rk}}, \tag{1.15}$$

并且,让原来的基变量 $x_{B_r}$ 变为非基变量,即令 $x_{B_r}=0$. 这样得到新的基本可行解

$$\hat{x}=(\hat{x}_{B_1},\cdots,\hat{x}_{B_{r-1}},0,\hat{x}_{B_{r+1}},\cdots,x_{B_m},0,\cdots,x_k,\cdots,0)^{\mathrm{T}}.$$

相应的目标函数值为

$$\hat{f}=c\hat{x}=\bar{f}-(z_k-c_k)x_k<\bar{f}.$$

且有

$$\hat{x}_{B_i}=\bar{b}_i-\bar{a}_{ik}x_k\geqslant 0\quad(i=1,\cdots,r-1,r+1,\cdots,m),$$

$$\hat{x}_{B_r}=x_k>0,$$

即,$\hat{\boldsymbol{x}}_B\geqslant\boldsymbol{0}$.

### 1.2.2 单纯形法的计算步骤

对标准形式的线性规划问题,单纯形法的计算步骤如下:

(1) 找初始可行基 $\boldsymbol{B}$ 和初始基本可行解.

(2) 求出 $\boldsymbol{x}_B=\boldsymbol{B}^{-1}\boldsymbol{b}\triangleq\bar{\boldsymbol{b}}$,计算目标函数值 $f=\boldsymbol{c}_B\boldsymbol{x}_B$.

(3) 按(1.10)计算检验数,并按

$$z_k-c_k=\max\{z_j-c_j\mid j=1,2,\cdots,n\},\qquad(1.16)$$

确定下标 $k$,取 $x_k$ 为进基变量.

(4) 若 $z_k-c_k\leqslant 0$,停止. 这时基本可行解

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\\boldsymbol{x}_N\end{bmatrix}=\begin{bmatrix}\bar{\boldsymbol{b}}\\\boldsymbol{0}\end{bmatrix}$$

是最优解,目标函数最优值为 $f=\boldsymbol{C}_B\bar{\boldsymbol{b}}$;否则,执行(5).

(5) 解 $\boldsymbol{B}\bar{\boldsymbol{a}}_k=\boldsymbol{a}_k$,求得 $\bar{\boldsymbol{a}}_k=\boldsymbol{B}^{-1}\boldsymbol{a}_k$. 若 $\bar{\boldsymbol{a}}_k\leqslant\boldsymbol{0}$,停止. 问题无界;否则,执行(6).

(6) 求最小比

$$\frac{\bar{b}_r}{\bar{a}_{rk}}=\min\left\{\frac{\bar{b}_i}{\bar{a}_{ik}}\,\middle|\,\bar{a}_{ik}>0\right\},$$

确定下标 $r$，$x_{B_r}$ 为离基变量.

(7) 以 $\boldsymbol{a}_k$ 代替 $\boldsymbol{a}_{B_r}$ 得到新基. 再返回(2).

### 1.2.3 单纯形表

本节介绍如何在表上实现单纯形法. 用单纯形法解线性规划时，主要是求改进的基本可行解. 为此，使原来的非基变量 $x_k$ 变成取正值的基变量. 同时，使原来的基变量 $x_{B_r}$ 取值为零，变成非基变量. 相应地，使原来的基 $\boldsymbol{B}=[\boldsymbol{a}_1\ \boldsymbol{a}_2\cdots\ \boldsymbol{a}_m]$ 变成新基 $\hat{\boldsymbol{B}}=[\boldsymbol{a}_1\cdots\boldsymbol{a}_{r-1}\ \boldsymbol{a}_k\ \boldsymbol{a}_{r+1}\cdots\ \boldsymbol{a}_m]$. 这种变换称为**换基**. 为了在表上实现换基运算，把在基 $\boldsymbol{B}$ 下式(1.13)的系数增广矩阵，式(1.12)中目标函数的系数及基本可行解中 $\boldsymbol{x}_B$ 各分量的值统列成下表，称为**单纯形表**(simplex tableau).

**表 1.1 单纯形表**

| | $x_1$ | $\cdots$ | $x_r$ | $\cdots$ | $x_m$ | $x_{m+1}$ | $\cdots$ | $x_k$ | $\cdots$ | $x_n$ | 右端 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $\cdots$ | 0 | $\cdots$ | 0 | $\bar{a}_{1,m+1}$ | $\cdots$ | $\bar{a}_{1k}$ | $\cdots$ | $\bar{a}_{1n}$ | $\bar{b}_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_r$ | 0 | $\cdots$ | 1 | $\cdots$ | 0 | $\bar{a}_{r,m+1}$ | $\cdots$ | $\bar{a}_{rk}$ | $\cdots$ | $\bar{a}_{rn}$ | $\bar{b}_r$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_m$ | 0 | $\cdots$ | 0 | $\cdots$ | 1 | $\bar{a}_{m,m+1}$ | $\cdots$ | $\bar{a}_{mk}$ | $\cdots$ | $\bar{a}_{mn}$ | $\bar{b}_m$ |
| $f$ | 0 | $\cdots$ | 0 | $\cdots$ | 0 | $z_{m+1}-c_{m+1}$ | $\cdots$ | $z_k-c_k$ | $\cdots$ | $z_n-c_n$ | $\bar{f}$ |

表中最右端一列对应于基本可行解中 $\boldsymbol{x}_B$ 各分量的值.

为了得到相应于新基 $\hat{\boldsymbol{B}}$ 的式(1.13)，就要把第 $k$ 列变成单位向量，其中第 $r$ 个元素为 1，其余元素为 0；从而第 $r$ 列不再是单位向量，必须按下列公式变换表中的元素：

$$\hat{a}_{rj}=\frac{\bar{a}_{rj}}{\bar{a}_{rk}}\ (j=1,2,\cdots,n). \tag{1.17}$$

$$\hat{a}_{ij} = \bar{a}_{ij} - \bar{a}_{ik}\,\hat{a}_{rj}$$

$$(j = 1,2,\cdots,n;\ i = 1,\cdots,r-1,r+1,\cdots,m). \quad (1.18)$$

$$\hat{b}_r = \frac{\bar{b}_r}{\bar{a}_{rk}}. \quad (1.19)$$

$$\hat{b}_i = \bar{b}_i - \hat{b}_r\bar{a}_{ik} \quad (i = 1,\cdots,r-1,r+1,\cdots,m). \quad (1.20)$$

得到对应于新基$\hat{B}$ 的单纯形表如表 1.2：

表 1.2

| | $x_1$ | $\cdots$ | $x_r$ | $\cdots$ | $x_m$ | $x_{m+1}$ | | $x_k$ | | $x_n$ | 右端 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $\cdots$ | $\hat{a}_{1r}$ | $\cdots$ | 0 | $\hat{a}_{1,m+1}$ | $\cdots$ | 0 | $\cdots$ | $\hat{a}_{1n}$ | $\hat{b}_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_k$ | 0 | $\cdots$ | $\hat{a}_{rr}$ | $\cdots$ | 0 | $\hat{a}_{r,m+1}$ | $\cdots$ | 1 | $\cdots$ | $\hat{a}_{rn}$ | $\hat{b}_r$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_m$ | 0 | $\cdots$ | $\hat{a}_{mr}$ | $\cdots$ | 1 | $\hat{a}_{m,m+1}$ | $\cdots$ | 0 | $\cdots$ | $\hat{a}_{mn}$ | $\hat{b}_m$ |
| $f$ | 0 | $\cdots$ | $\hat{z}_r - c_r$ | $\cdots$ | 0 | $\hat{z}_{m+1} - c_{m+1}$ | $\cdots$ | 0 | $\cdots$ | $\hat{z}_n - c_n$ | $\hat{f}$ |

迭代时，根据定理 1.2.3 确定 $x_k$ 为进基变量，$x_{B_r}$ 为离基变量，$\bar{a}_{rk}$为主元. 然后，进行换基，即按(1.17)～(1.20)变换表中的元素. 这样，单纯形法全部计算都可以在单纯形表上进行了.

单纯形表也可简记为

表 1.3

| | $\boldsymbol{x}_B$ | $\boldsymbol{x}_N$ | 右端 |
|---|---|---|---|
| $\boldsymbol{x}_B$ | $\boldsymbol{I}$ | $\boldsymbol{B}^{-1}N$ | $\bar{\boldsymbol{b}}$ |
| $f$ | $\boldsymbol{0}$ | $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{N}-\boldsymbol{c}_N$ | $\boldsymbol{c}_B\bar{\boldsymbol{b}}$ |

**例 1.2.4** 用单纯形法求解下列线性规划

$$
\begin{aligned}
\min \quad & x_1 + x_2 - 4x_3; \\
\text{s.t.} \quad & x_1 + x_2 + 2x_3 + x_4 = 9, \\
& x_1 + x_2 - x_3 + x_5 = 2, \\
& -x_1 + x_2 + x_3 + x_6 = 4, \\
& x_j \geqslant 0\ (j = 1,2,\cdots,6).
\end{aligned}
$$

**解** 由于$[\boldsymbol{a}_4\ \boldsymbol{a}_5\ \boldsymbol{a}_6]$是一个单位矩阵，且$\boldsymbol{b}=(9,2,4)^{\mathrm{T}}\geqslant\boldsymbol{0}$，故选取初始基为$\boldsymbol{B}=[\boldsymbol{a}_4\ \boldsymbol{a}_5\ \boldsymbol{a}_6]$，于是有$\boldsymbol{B}^{-1}\boldsymbol{b}=\bar{\boldsymbol{b}}\geqslant\boldsymbol{0}$. 这样，得到初始单纯形表（表 1.4）如下：

**迭代 1**

表 1.4

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_4$ | 1 | 1 | 2 | 1 | 0 | 0 | 9 |
| $x_5$ | 1 | 1 | −1 | 0 | 1 | 0 | 2 |
| $x_6$ | −1 | 1 | [1] | 0 | 0 | 1 | 4 |
| $f$ | −1 | −1 | 4 | 0 | 0 | 0 | 0 |

对应的基本可行解为$\boldsymbol{x}=(0,0,0,9,2,4)^{\mathrm{T}}$，目标函数值为$f=0$. 因检验数$z_3-c_3=4>0$，故不是最优解. 因为检验数中只有$z_3-c_3$是正的，故主列是$\bar{a}_3$，相应的$x_3$为进基变量. 在主列中$\bar{a}_{13}$和$\bar{a}_{33}$是正的，并且有$\bar{b}_1/\bar{a}_{13}=9/2>\bar{b}_3/\bar{a}_{33}=4$，故主元为$\bar{a}_{33}$. 相应地，$x_6$为离基变量. 以$\bar{a}_{33}$为主元，通过消元法运算得表 1.5：

**迭代 2**

表 1.5

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_4$ | $\boxed{3}$ | $-1$ | 0 | 1 | 0 | $-2$ | 1 |
| $x_5$ | 0 | 2 | 0 | 0 | 1 | 1 | 6 |
| $x_3$ | $-1$ | 1 | 1 | 0 | 0 | 1 | 4 |
| $f$ | 3 | $-5$ | 0 | 0 | 0 | $-4$ | $-16$ |

它所对应的基本可行解是 $\boldsymbol{x}=(0,0,4,1,6,0)^{\mathrm{T}}$，目标函数值 $f=-16$，比迭代 1 的解有改进. 但是，$z_1-c_1=3>0$，故仍然不是最优解. 以 $\bar{a}_{11}$ 为主元，通过消元法运算得表 1.6：

**迭代 3**

表 1.6

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $-\frac{1}{3}$ | 0 | $\frac{1}{3}$ | 0 | $-\frac{2}{3}$ | $\frac{1}{3}$ |
| $x_5$ | 0 | 2 | 0 | 0 | 1 | 1 | 6 |
| $x_3$ | 0 | $\frac{2}{3}$ | 1 | $\frac{1}{3}$ | 0 | $\frac{1}{3}$ | $\frac{13}{3}$ |
| $f$ | 0 | $-4$ | 0 | $-1$ | 0 | $-2$ | $-17$ |

它所对应的基本可行解是 $\boldsymbol{x}=\left(\frac{1}{3},0,\frac{13}{3},0,6,0\right)^{\mathrm{T}}$，目标函数值 $f=-17$. 由于这时所有的检验数 $z_j-c_j\leqslant 0$，故为最优解.

### 1.2.4 两阶段法

用单纯形法解线性规划问题时，需要先有一个初始基本可行解. 如果一个线性规划问题像例 1.2.4 那样，在约束矩阵 $\boldsymbol{A}$ 中含有一个单位矩阵，且 $\boldsymbol{b}\geqslant\mathbf{0}$. 似乎取 $\boldsymbol{B}=\boldsymbol{I}$，即得一个基本可行解. 实

际上并非如此简单，因此还需要一些寻找初始基本可行解的方法.下面介绍**两阶段法**(two-phase method).

第一阶段引入人工变量，求初始基本可行解；第二阶段从初始基本可行解开始，用单纯形法求解原问题.详述如下：

考虑(1.3)，设 $\boldsymbol{A}$ 中不含单位矩阵.引入人工变量 $\boldsymbol{x}_a=(x_{n+1},x_{n+2},\cdots,x_{n+m})^{\mathrm{T}}$，构造辅助问题：

$$\begin{aligned} \min \quad & g = \boldsymbol{e}\boldsymbol{x}_a; \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}_a = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{x}_a \geqslant \boldsymbol{0}. \end{aligned} \tag{1.21}$$

其中，$\boldsymbol{e}=(1,1,\cdots,1)$是分量全等于1的 $m$ 维行向量.

由于在(1.21)中人工变量对应的 $m$ 列构成单位矩阵，且 $\boldsymbol{b}\geqslant\boldsymbol{0}$.所以，辅助问题有初始基本可行解$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{x}_a\end{bmatrix}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}\end{bmatrix}$，相应的目标函数值 $g=\sum_{i=1}^{m}b_i$.于是可由此基本可行解开始进行单纯形法迭代.注意到 $g\geqslant0$，故辅助问题必有最优解，设其最优解为$\begin{bmatrix}\bar{\boldsymbol{x}}\\ \bar{\boldsymbol{x}}_a\end{bmatrix}$，并且只可能出现如下3种情形：

(1) $\min g>0$.这说明不存在 $\boldsymbol{x}$ 使$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{0}\end{bmatrix}$满足(1.21).即，原问题无可行解.

(2) $\min g=0$，即 $\bar{\boldsymbol{x}}_a=\boldsymbol{0}$ 且 $\bar{\boldsymbol{x}}_a$ 的分量都是非基变量.这时基变量全是原问题的变量，又知$\begin{bmatrix}\bar{\boldsymbol{x}}\\ \boldsymbol{0}\end{bmatrix}$是(1.21)的基本可行解，所以，$\boldsymbol{x}=\bar{\boldsymbol{x}}$ 是原问题的一个基本可行解.故可以由此开始求解原问题.

(3) $\min g=0$，且 $\bar{\boldsymbol{x}}_a$ 的某些分量为基变量.这时，可用消元法将含在基变量中的人工变量换出来.设辅助问题的最优单纯形表如下：

表 1.7

| | $x_1$ | $\cdots$ | $x_s$ | $\cdots$ | $x_n$ | $x_{n+1}$ | $\cdots$ | $x_{n+m}$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|---|
| $x_{B_1}$ | $\bar{a}_{11}$ | $\cdots$ | $\bar{a}_{1s}$ | $\cdots$ | $\bar{a}_{1n}$ | $\bar{a}_{1,n+1}$ | $\cdots$ | $\bar{a}_{1,n+m}$ | $\bar{b}_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_{B_r}$ | $\bar{a}_{r1}$ | $\cdots$ | $\bar{a}_{rs}$ | $\cdots$ | $\bar{a}_{rn}$ | $\bar{a}_{r,n+1}$ | $\cdots$ | $\bar{a}_{r,n+m}$ | $\bar{b}_r$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_{B_m}$ | $\bar{a}_{m1}$ | $\cdots$ | $\bar{a}_{ms}$ | $\cdots$ | $\bar{a}_{mn}$ | $\bar{a}_{m,n+1}$ | $\cdots$ | $\bar{a}_{m,n+m}$ | $\bar{b}_m$ |
| $f$ | $y_1$ | | $y_s$ | | $y_n$ | $y_{n+1}$ | | $y_{n+m}$ | 0 |

其中，$y_1,\cdots,y_s,\cdots,y_n,y_{n+1},\cdots,y_{n+m}$为检验数，基变量是 $x_{B_1},\cdots,x_{B_r},\cdots,x_{B_m}$. 假设其中 $x_{B_r}$是一个人工变量.

若第 $r$ 行的前 $n$ 个元素不全为零，设 $\bar{a}_{rs}\neq 0(1\leqslant s\leqslant n)$，则以$\bar{a}_{rs}$为主元进行消元法运算. 由于 $x_{B_r}$是人工变量，故 $\bar{b}_r=0$. 因此，经过一次消元法运算以后，目标函数最优值不变化. 但是，将 $x_s$ 变成了基变量而换出了人工变量 $x_{B_r}$. 若在基变量中还有人工变量，都按此方法换出来，最终可以求得原问题的一个基本可行解，便可由此开始对原问题进行单纯形迭代.

如果第 $r$ 行的前 $n$ 个元素全为零，于是矩阵

$$\bar{\boldsymbol{A}}=\begin{bmatrix}\bar{a}_{11} & \cdots & \bar{a}_{1n}\\ \vdots & & \vdots\\ \bar{a}_{m1} & \cdots & \bar{a}_{mn}\end{bmatrix}$$

的秩 $\mathrm{rank}(\bar{\boldsymbol{A}})<m$，从而矩阵 $\boldsymbol{A}$ 的秩也小于 $m$. 这说明第 $r$ 个约束是多余的，把它删去.

**例 1.2.5** 用两阶段法求解下列线性规划

$$\min \quad f=-2x_1-x_2;$$

$$\text{s.t.}\quad \begin{aligned} x_1 + x_2 - x_3 \qquad &= 3, \\ -x_1 + x_2 \qquad - x_4 \qquad &= 2, \\ x_2 \qquad + x_5 &= 3, \end{aligned}$$

$$x_1, x_2, x_3, x_4, x_5 \geqslant 0.$$

**解** 由于没有明显的基本可行解,故引入人工变量 $x_6, x_7$,写出辅助问题

$$\min \quad g = x_6 + x_7;$$

$$\text{s.t.}\quad \begin{aligned} x_1 + x_2 - x_3 \qquad + x_6 \qquad &= 3, \\ -x_1 + x_2 \qquad - x_4 \qquad + x_7 &= 2, \\ x_2 \qquad + x_5 \qquad &= 3, \end{aligned}$$

$$x_j \geqslant 0 \quad (j = 1, 2, \cdots, 7).$$

用单纯形法解此辅助问题,迭代过程单纯形表 1.8 如下:

**表 1.8**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | 1 | 1 | −1 | 0 | 0 | 1 | 0 | 3 |
| $x_7$ | −1 | [1] | 0 | −1 | 0 | 0 | 1 | 2 |
| $x_5$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 |
| $g$ | 0 | 2 | −1 | −1 | 0 | 0 | 0 | 5 |

**表 1.9**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | [2] | 0 | −1 | 1 | 0 | 1 | −1 | 1 |
| $x_2$ | −1 | 1 | 0 | −1 | 0 | 0 | 1 | 2 |
| $x_5$ | 1 | 0 | 0 | 1 | 1 | 0 | −1 | 1 |
| $g$ | 2 | 0 | −1 | 1 | 0 | 0 | −2 | 1 |

表 1.10

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | $-\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{1}{2}$ |
| $x_2$ | 0 | 1 | $-\frac{1}{2}$ | $-\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{5}{2}$ |
| $x_5$ | 0 | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 | $-\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{1}{2}$ |
| $g$ | 0 | 0 | 0 | 0 | 0 | $-1$ | $-1$ | 0 |

最后的表 1.10 对应的基变量是 $x_1, x_2, x_5$. 基本可行解是 $\left(\frac{1}{2},\frac{5}{2},0,0,\frac{1}{2},0,0\right)^{\mathrm{T}}$,辅助问题的目标函数最优值 $g=0$,故第一阶段结束. 删去 $x_6, x_7$ 对应的两列,按照原问题目标函数的系数修改检验行,并从基本可行解 $\boldsymbol{x}=\left(\frac{1}{2},\frac{5}{2},0,0,\frac{1}{2},0,0\right)^{\mathrm{T}}$ 开始第二阶段迭代. 第二阶段是极小化原问题的目标函数 $f=-2x_1-x_2$. 迭代过程如表 1.11:

表 1.11

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | $-\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| $x_2$ | 0 | 1 | $-\frac{1}{2}$ | $-\frac{1}{2}$ | 0 | $\frac{5}{2}$ |
| $x_5$ | 0 | 0 | $\boxed{\frac{1}{2}}$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ |
| $f$ | 0 | 0 | $\frac{3}{2}$ | $-\frac{1}{2}$ | 0 | $-\frac{7}{2}$ |

表 1.12

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 1 | 1 | 1 |
| $x_2$ | 0 | 1 | 0 | 0 | 1 | 3 |
| $x_3$ | 0 | 0 | 1 | 1 | 2 | 1 |
| $f$ | 0 | 0 | 0 | −2 | −3 | −5 |

在表 1.12 中，所有的检验数 $z_j-c_j\leqslant 0$，故已得到原问题的最优解为

$$\boldsymbol{x}=(1,3,1,0,0)^{\mathrm{T}}.$$

目标函数最优值为　$f=-5$.

### 1.2.5　大 $M$ 法

在没有明显的基本可行解时，还可以采用**大 $M$ 法**(big-$M$ method)求解线性规划问题.

考虑(1.3)，像两阶段法一样，在约束中引入人工变量 $\boldsymbol{x}_a$，并且在目标函数中加上惩罚项 $M\boldsymbol{e}\boldsymbol{x}_a$，得如下问题：

$$\begin{aligned}\min\quad & \boldsymbol{cx}+M\boldsymbol{ex}_a;\\ \text{s.t.}\quad & \boldsymbol{Ax}+\boldsymbol{x}_a=\boldsymbol{b},\\ & \boldsymbol{x}\geqslant\boldsymbol{0},\ \boldsymbol{x}_a\geqslant\boldsymbol{0}.\end{aligned}\tag{1.22}$$

其中，$M$ 是足够大的正数. 由于惩罚项 $M\boldsymbol{ex}_a$ 的存在，在极小化目标函数值的迭代过程中，将迫使人工变量 $\boldsymbol{x}_a=\boldsymbol{0}$. 这就是说，$\bar{\boldsymbol{x}}$ 是(1.3)的最优解与 $\begin{bmatrix}\bar{\boldsymbol{x}}\\ \boldsymbol{0}\end{bmatrix}$ 是(1.22)的最优解是等价的. 由于(1.22)有初始基本可行解 $\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{b}\end{bmatrix}$，因此可用单纯形法求解.

**例 1.2.6**　用大 $M$ 法求解下列线性规划

$$\min\quad f=-2x_1-x_2;$$

$$\text{s.t.}\quad \begin{aligned} x_1 + x_2 &\geqslant 3, \\ -x_1 + x_2 &\geqslant 2, \\ x_2 &\leqslant 3, \\ x_1 \geqslant 0,\quad x_2 &\geqslant 0. \end{aligned}$$

**解** 引入松弛变量 $x_3, x_4$ 和 $x_5$，再引入人工变量 $x_6, x_7$，于是问题变换为

$$\min \quad -2x_1 - x_2 + M(x_6 + x_7);$$

$$\text{s.t.}\quad \begin{aligned} x_1 + x_2 - x_3 + x_6 &= 3, \\ -x_1 + x_2 - x_4 + x_7 &= 2, \\ x_2 + x_5 &= 3, \\ x_j \geqslant 0 \quad j = (1,2,\cdots,7)&. \end{aligned}$$

用单纯形法解此问题，迭代过程如表 1.13：

**表 1.13**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | 1 | 1 | $-1$ | 0 | 0 | 1 | 0 | 3 |
| $x_7$ | $-1$ | $\boxed{1}$ | 0 | $-1$ | 0 | 0 | 1 | 2 |
| $x_5$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 |
| $f$ | 2 | $2M+1$ | $-M$ | $-M$ | 0 | 0 | 0 | $5M$ |

注意到 $M$ 是足够大的正数，故取 $x_2$ 为进基变量，人工变量 $x_7$ 离基，继续迭代.

**表 1.14**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | $\boxed{2}$ | 0 | $-1$ | 1 | 0 | 1 | $-1$ | 1 |
| $x_2$ | $-1$ | 1 | 0 | $-1$ | 0 | 0 | 1 | 2 |
| $x_5$ | 1 | 0 | 0 | 1 | 1 | 0 | $-1$ | 1 |
| $f$ | $2M+3$ | 0 | $-M$ | $M+1$ | 0 | 0 | $-2M-1$ | $M-2$ |

表 1.15

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | $-\frac{1}{2}$ | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{1}{2}$ |
| $x_2$ | 0 | 1 | $-\frac{1}{2}$ | $-\frac{1}{2}$ | 0 | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{5}{2}$ |
| $x_5$ | 0 | 0 | $\boxed{\frac{1}{2}}$ | $\frac{1}{2}$ | 1 | $-\frac{1}{2}$ | $-\frac{1}{2}$ | $\frac{1}{2}$ |
| $f$ | 0 | 0 | $\frac{3}{2}$ | $-\frac{1}{2}$ | 0 | $-M-\frac{3}{2}$ | $-M+\frac{1}{2}$ | $-\frac{7}{2}$ |

表 1.16

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 1 | 1 | 0 | $-1$ | 1 |
| $x_2$ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 |
| $x_3$ | 0 | 0 | 1 | 1 | 2 | $-1$ | $-1$ | 1 |
| $f$ | 0 | 0 | 0 | $-2$ | $-3$ | $-M$ | $-M+2$ | $-5$ |

由于 $M$ 是足够大的正数,故所有的检验数 $z_j-c_j\leqslant 0$. 这样就得到了原问题的最优解 $\boldsymbol{x}=(1,3,1,0,0)^{\mathrm{T}}$,目标函数最优值 $f=-5$,与用两阶段法解此问题的结果相同.

### 1.2.6 退化与防止循环

如果在基本可行解中,有一个或几个基变量为零,则称为退化的基本可行解.

在基本可行解退化时,有可能发生(虽然可能性极小)用单纯形法要进行无限多次迭代,也不能得到最优解的情形.

**例 1.2.7** 用单纯形法解下列问题

$$
\begin{aligned}
\min \quad & -\frac{3}{4}x_4 + 20x_5 - \frac{1}{2}x_6 + 6x_7; \\
\text{s.t.} \quad & x_1 + \frac{1}{4}x_4 - 8x_5 - x_6 + 9x_7 = 0, \\
& x_2 + \frac{1}{2}x_4 - 12x_5 - \frac{1}{2}x_6 + 3x_7 = 0, \\
& x_3 + x_6 = 1, \\
& x_j \geqslant 0 \quad j = (1,2,\cdots,7).
\end{aligned}
$$

**解** 用单纯形法解这个问题,迭代过程如下:

**表 1.17**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | $\boxed{\frac{1}{4}}$ | $-8$ | $-1$ | 9 | 0 |
| $x_2$ | 0 | 1 | 0 | $\frac{1}{2}$ | $-12$ | $-\frac{1}{2}$ | 3 | 0 |
| $x_7$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | 0 | 0 | 0 | $\frac{3}{4}$ | $-20$ | $\frac{1}{2}$ | $-6$ | 0 |

**表 1.18**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_4$ | 4 | 0 | 0 | 1 | $-32$ | $-4$ | 36 | 0 |
| $x_2$ | $-2$ | 1 | 0 | 0 | $\boxed{4}$ | $\frac{3}{2}$ | $-15$ | 0 |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | $-3$ | 0 | 0 | 0 | 4 | $\frac{7}{2}$ | $-33$ | 0 |

表 1.19

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_4$ | $-12$ | $8$ | $0$ | $1$ | $0$ | $\boxed{8}$ | $-84$ | $0$ |
| $x_5$ | $-\frac{1}{2}$ | $\frac{1}{4}$ | $0$ | $0$ | $1$ | $\frac{3}{8}$ | $-\frac{15}{4}$ | $0$ |
| $x_3$ | $0$ | $0$ | $1$ | $0$ | $0$ | $1$ | $0$ | $1$ |
| $f$ | $-1$ | $-1$ | $0$ | $0$ | $0$ | $2$ | $-18$ | $0$ |

表 1.20

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | $-\frac{3}{2}$ | $1$ | $0$ | $\frac{1}{8}$ | $0$ | $1$ | $-\frac{21}{2}$ | $0$ |
| $x_5$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $0$ | $-\frac{3}{64}$ | $1$ | $0$ | $\boxed{\frac{3}{16}}$ | $0$ |
| $x_3$ | $\frac{3}{2}$ | $-1$ | $1$ | $-\frac{1}{8}$ | $0$ | $0$ | $\frac{21}{2}$ | $1$ |
| $f$ | $2$ | $-3$ | $0$ | $-\frac{1}{4}$ | $0$ | $0$ | $3$ | $0$ |

表 1.21

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_6$ | $\boxed{2}$ | $-6$ | $0$ | $-\frac{5}{2}$ | $56$ | $1$ | $0$ | $0$ |
| $x_7$ | $\frac{1}{3}$ | $-\frac{2}{3}$ | $0$ | $-\frac{1}{4}$ | $\frac{16}{3}$ | $0$ | $1$ | $0$ |
| $x_3$ | $-2$ | $6$ | $1$ | $\frac{5}{2}$ | $-56$ | $0$ | $0$ | $1$ |
| $f$ | $1$ | $-1$ | $0$ | $\frac{1}{2}$ | $-16$ | $0$ | $0$ | $0$ |

表 1.22

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $-3$ | 0 | $-\frac{5}{4}$ | 28 | $\frac{1}{2}$ | 0 | 0 |
| $x_7$ | 0 | $\boxed{\frac{1}{3}}$ | 0 | $\frac{1}{6}$ | $-4$ | $-\frac{1}{6}$ | 1 | 0 |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | 0 | 2 | 0 | $\frac{7}{4}$ | $-44$ | $-\frac{1}{2}$ | 0 | 0 |

表 1.23

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | $\frac{1}{4}$ | $-8$ | $-1$ | 9 | 0 |
| $x_2$ | 0 | 1 | 0 | $\frac{1}{2}$ | $-12$ | $-\frac{1}{2}$ | 3 | 0 |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | 0 | 0 | 0 | $\frac{3}{4}$ | $-20$ | $\frac{1}{2}$ | $-6$ | 0 |

经过 6 次迭代以后，得到的表 1.23 与初始单纯形表 1.17 相同. 因此，如果按照原来的规则再迭代下去，必将得到上面这些表的重复，从而永远得不到最优解（实际上，该问题的最优解是存在的）. 这时，就出现**循环**(cycling).

现介绍一种避免循环的方法，称为**字典序规则**(lexicographic rule).

**定义 1.2.8** 若向量 $\boldsymbol{x}\neq\mathbf{0}$，且它的第一个非零分量是正的，则称 $\boldsymbol{x}$ 是**按字典序正的**(lexicographically positive). 记成 $\boldsymbol{x}\succ\mathbf{0}$；若 $\boldsymbol{x}=\mathbf{0}$ 或是按字典序正的，则称 $\boldsymbol{x}$ 是**按字典序非负的**. 记成 $\boldsymbol{x}\succcurlyeq\mathbf{0}$.

**定义 1.2.9** 若对向量 $x$ 和 $y$，有 $x-y\succ 0$，则称 $x$ **按字典序大于** $y$. 记成 $x\succ y$；若有 $x-y\succeq 0$，则称 $x\succeq y$.

**定义 1.2.10** 若在一组 $x^{(i)}$ 中，有 $x^{(t)}$ 对所有的 $i$ 均有 $x^{(i)}\succeq x^{(t)}$，则称 $x^{(t)}$ 为这组向量中**按字典序最小的**. 记成 $x^{(t)}=\operatorname{lex}\min\limits_i x^{(i)}$.

**例 1.2.11** $x=(0,0,3,-1,2)$，$y=(0,2,4,0,3)$ 都是按字典序正的. 并且，$y$ 按字典序大于 $x$.

字典序规则的特点如下：

考虑(1.3)，给定初始基本可行解，建立单纯形表

**表 1.24**

| | $x_B$ | $x_N$ | 右端 |
|---|---|---|---|
| $x_B$ | $I$ | $B^{-1}N$ | $\bar{b}$ |
| $f$ | $0$ | $c_BB^{-1}N-c_N$ | $f_0$ |

令 $P_i$ 表示$(\bar{b}\ B^{-1})$的第 $i$ 个行向量.

字典序规则是在确定了进基变量 $x_k$ 以后，令

$$\frac{p_r}{\bar{a}_{rk}}=\operatorname{lex}\min_i\left\{\frac{p_i}{\bar{a}_{ik}}\,\middle|\,\bar{a}_{ik}>0\right\}.$$

确定了 $r$ 以后，选 $x_{B_r}$ 为离基变量. 然后，以 $\bar{a}_{rk}$ 为主元进行消元法运算.

**定理 1.2.12** 按照字典序规则进行迭代，单纯形法不会出现循环.

**例 1.2.13** 按字典序规则迭代，求解例 1.2.7.

**解** 建立初始单纯形表 1.25：

表 1.25

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | $\frac{1}{4}$ | $-8$ | $-1$ | 9 | 0 |
| $x_2$ | 0 | 1 | 0 | $\boxed{\frac{1}{2}}$ | $-12$ | $-\frac{1}{2}$ | 3 | 0 |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | 0 | 0 | 0 | $\frac{3}{4}$ | $-20$ | $\frac{1}{2}$ | $-6$ | 0 |

这里，$\boldsymbol{P}_1=(0,1,0,0)$，$\boldsymbol{P}_2=(0,0,1,0)$，$\boldsymbol{P}_3=(1,0,0,1)$. $x_4$ 为进基变量，$\bar{a}_{14}=\frac{1}{4}>0$，$\bar{a}_{24}=\frac{1}{2}>0$. 于是

$$\frac{\boldsymbol{P}_1}{\bar{a}_{14}}=(0,4,0,0),\ \frac{\boldsymbol{P}_2}{\bar{a}_{24}}=(0,0,2,0),$$

故$\frac{p_2}{\bar{a}_{24}}=\operatorname{lex}\min\limits_i\left\{\frac{\boldsymbol{P}_i}{\bar{a}_{i4}}\middle|\bar{a}_{i4}>0\right\}$. 所以，$x_{B_2}=x_2$ 为离基变量(注意，在例 1.2.7 的第一次迭代时是 $x_1$ 离基). 以 $\bar{a}_{24}$为主元进行消元法得表 1.26:

表 1.26

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $-\frac{1}{2}$ | 0 | 0 | $-2$ | $-\frac{3}{4}$ | $\frac{15}{2}$ | 0 |
| $x_4$ | 0 | 2 | 0 | 1 | $-24$ | $-1$ | 6 | 0 |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | $\boxed{1}$ | 0 | 1 |
| $f$ | 0 | $-\frac{3}{2}$ | 0 | 0 | $-2$ | $\frac{5}{4}$ | $-\frac{21}{2}$ | 0 |

现在，$x_6$ 进基，$x_3$ 离基，以 $\bar{a}_{36}$为主元进行消元法得表 1.27

表 1.27

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $-\frac{1}{2}$ | $\frac{3}{4}$ | 0 | $-2$ | 0 | $\frac{15}{2}$ | $\frac{3}{4}$ |
| $x_4$ | 0 | 2 | 1 | 1 | $-24$ | 0 | 6 | 1 |
| $x_6$ | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $f$ | 0 | $-\frac{3}{2}$ | $-\frac{5}{4}$ | 0 | $-2$ | 0 | $-\frac{21}{2}$ | $-\frac{5}{4}$ |

由于所有的检验数 $z_j-c_j\leqslant 0$，故表 1.27 已经给出最优解，计算停止. 最优解为 $x=\left(\frac{3}{4},0,0,1,0,1,0\right)^T$，目标函数最优值 $f_0=-\frac{5}{4}$.

## 1.3 修正单纯形法

### 1.3.1 修正单纯形法原理

一般，实际中的线性规划问题规模较大，即变量的个数和约束的个数较多. 因此，使用计算机求解时必须节省存储单元和减少计算时间. 修正单纯形法(revised simplex method)就是解决这些问题的一种实用、有效的方法.

修正单纯形法只要存储下列所谓修正单纯形表中的数据和原始数据.

**表 1.28 修正单纯形表**

| $\boldsymbol{w}=\boldsymbol{c}_B\boldsymbol{B}^{-1}$ | $f=\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$ |
|---|---|
| $\boldsymbol{B}^{-1}$ | $\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\boldsymbol{b}$ |

每次迭代时，仍然是根据(1.10)，(1.16) $z_k-c_k=\max\limits_{j\in N_B}\{\boldsymbol{w}\boldsymbol{a}_j-c_j$

$|N_B$ 为非基变量的下标集$\}$确定 $x_k$ 为进基变量，而用最小比确定离基变量 $x_{B_r}$. 按下法修改旧基逆 $\boldsymbol{B}^{-1}$ 为新基逆$\hat{\boldsymbol{B}}^{-1}$：

在对应于基 $\boldsymbol{B}$ 的修正单纯形表的右端添加新的一列

| $z_k-c_k$ |
|---|
| $\bar{a}_{1k}$<br>$\vdots$<br>$\bar{a}_{rk}$<br>$\vdots$<br>$\bar{a}_{mk}$ |

以 $\bar{a}_{rk}$为主元进行消元法运算，就得到对应于基$\hat{\boldsymbol{B}}$ 的修正单纯形表.

### 1.3.2 修正单纯形法的计算步骤

(1) 找初始可行基 $\boldsymbol{B}$，计算 $\boldsymbol{B}^{-1}$，再计算 $\boldsymbol{w}=\boldsymbol{c}_B\boldsymbol{B}^{-1}$，$\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\boldsymbol{b}$，$f=\boldsymbol{c}_B\bar{\boldsymbol{b}}$，构造初始修正单纯形表

| | | |
|---|---|---|
| | $\boldsymbol{w}$ | $f$ |
| $\boldsymbol{x}_B$ | $\boldsymbol{B}^{-1}$ | $\bar{\boldsymbol{b}}$ |

(2) 求 $z_k-c_k=\max\limits_{j\in N_B}\{\boldsymbol{w}\boldsymbol{a}_j-c_j|N_B$ 为非基变量的下标集$\}$. 若 $z_k-c_k\leqslant 0$，停止. 已求得最优解 $\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\\boldsymbol{x}_N\end{bmatrix}=\begin{bmatrix}\bar{\boldsymbol{b}}\\\boldsymbol{0}\end{bmatrix}$及目标函数最优值 $f$；否则，执行(3).

(3) 计算 $\bar{\boldsymbol{a}}_k=\boldsymbol{B}^{-1}\boldsymbol{a}_k$. 若 $\bar{\boldsymbol{a}}_k\leqslant 0$，停止. 问题无解；否则，执行(4).

(4) 在表的右侧添加新的一列 $\begin{bmatrix}z_k-c_k\\\bar{\boldsymbol{a}}_k\end{bmatrix}$，求 $\dfrac{\bar{b}_r}{\bar{a}_{rk}}=$

$\min\left\{\dfrac{\bar{b}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik}>0\right\}$.

(5) 以$\bar{a}_{rk}$为主元，进行消元法运算，然后返回(2).

**例 1.3.1** 用修正单纯形法解下列问题

$$
\begin{aligned}
\min \quad & f = -2x_2 + x_3; \\
\text{s.t.} \quad & x_1 - 2x_2 + x_3 \geqslant -4, \\
& x_1 + x_2 + x_3 \leqslant 9, \\
& 2x_1 - x_2 - x_3 \leqslant 5, \\
& x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0.
\end{aligned}
$$

**解** 先将第 1 个约束不等式两端同乘以(−1)，再引入松弛变量 $x_4$，$x_5$ 和 $x_6$，将问题化成标准形式

$$
\begin{aligned}
\min \quad & -2x_2 + x_3; \\
\text{s.t.} \quad & -x_1 + 2x_2 - x_3 + x_4 = 4, \\
& x_1 + x_2 + x_3 + x_5 = 9, \\
& 2x_1 - x_2 - x_3 + x_6 = 5, \\
& x_j \geqslant 0 \quad (j = 1,2,\cdots,6).
\end{aligned}
$$

约束方程的系数矩阵为

$$
\boldsymbol{A} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3 \quad \boldsymbol{a}_4 \quad \boldsymbol{a}_5 \quad \boldsymbol{a}_6]
= \begin{bmatrix} -1 & 2 & -1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & -1 & -1 & 0 & 0 & 1 \end{bmatrix}.
$$

约束方程的右端向量为

$$\boldsymbol{b} = (4,9,5)^{\mathrm{T}}.$$

它有一个初始基 $\boldsymbol{B}=[\boldsymbol{a}_4 \quad \boldsymbol{a}_5 \quad \boldsymbol{a}_6]=\boldsymbol{I}_3$. 按定义计算 $\boldsymbol{w}=\boldsymbol{c}_B\boldsymbol{B}^{-1}=(0,0,0)$，$\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\boldsymbol{b}=(4,9,5)^{\mathrm{T}}$，$f=\boldsymbol{c}_B\bar{\boldsymbol{b}}=0$，从而可得初始修正单纯形表 1.29

表 1.29

| | | | | |
|---|---|---|---|---|
| | 0 | 0 | 0 | 0 |
| $x_4$ | 1 | 0 | 0 | 4 |
| $x_5$ | 0 | 1 | 0 | 9 |
| $x_6$ | 0 | 0 | 1 | 5 |

**迭代 1**

$$\begin{aligned}&(z_1-c_1,z_2-c_2,z_3-c_3)\\&=\boldsymbol{w}[\boldsymbol{a}_1\quad\boldsymbol{a}_2\quad\boldsymbol{a}_3]-(c_1,c_2,c_3)\\&=(0,0,0)\begin{bmatrix}-1&2&-1\\1&1&1\\2&-1&-1\end{bmatrix}-(0,-2,1)\\&=(0,2,-1)\end{aligned}$$

因 $z_2-c_2=\max\{z_j-c_j\}$，故 $x_2$ 为进基变量. 计算主列

$$\bar{\boldsymbol{a}}_2=\boldsymbol{B}^{-1}\boldsymbol{a}_2=\begin{bmatrix}2\\1\\-1\end{bmatrix}$$

由于 $\bar{\boldsymbol{a}}_2$ 含有正分量，将$\begin{bmatrix}z_2-c_2\\\bar{\boldsymbol{a}}_2\end{bmatrix}$放在表 1.29 的右面，构造下表.

表 1.30

| | | | | | |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 2 |
| $x_4$ | 1 | 0 | 0 | 4 | [2] |
| $x_5$ | 0 | 1 | 0 | 9 | 1 |
| $x_6$ | 0 | 0 | 1 | 5 | −1 |

按最小比原则确定 $x_4$ 为离基变量，以 $\bar{a}_{12}=2$ 为主元进行消元法运

算得到下表 1.31

表 1.31

| | | | | |
|---|---|---|---|---|
| | $-1$ | 0 | 0 | $-4$ |
| $x_2$ | $\frac{1}{2}$ | 0 | 0 | 2 |
| $x_5$ | $-\frac{1}{2}$ | 1 | 0 | 7 |
| $x_6$ | $\frac{1}{2}$ | 0 | 1 | 7 |

**迭代 2**

**计算检验数得**

$$\begin{aligned}&(z_1 - c_1, z_3 - c_3, z_4 - c_4)\\&= \boldsymbol{w}[\boldsymbol{a}_1 \quad \boldsymbol{a}_3 \quad \boldsymbol{a}_4] - (c_1, c_3, c_4)\\&= (1, 0, -1).\end{aligned}$$

因 $z_1 - c_1 = \max\{z_j - c_j\}$，故 $x_1$ 为进基变量. 计算主列

$$\bar{\boldsymbol{a}}_1 = \boldsymbol{B}^{-1}\boldsymbol{a}_1 = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}.$$

将$\begin{bmatrix} z_1 - c_1 \\ \bar{\boldsymbol{a}}_1 \end{bmatrix}$放在表 1.31 的右面，得表 1.32

表　1.32

| | | | | | |
|---|---|---|---|---|---|
| | $-1$ | 0 | 0 | $-4$ | 1 |
| $x_2$ | $\frac{1}{2}$ | 0 | 0 | 2 | $-\frac{1}{2}$ |
| $x_5$ | $-\frac{1}{2}$ | 1 | 0 | 7 | $\boxed{\frac{3}{2}}$ |
| $x_6$ | $\frac{1}{2}$ | 0 | 1 | 7 | $\frac{3}{2}$ |

确定 $x_5$ 为离基变量，以 $\bar{a}_{21}=3/2$ 为主元进行消元法运算得到表 1.33.

表　1.33

| | | | | |
|---|---|---|---|---|
| | $-\frac{2}{3}$ | $-\frac{2}{3}$ | 0 | $-\frac{26}{3}$ |
| $x_2$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 0 | $\frac{13}{3}$ |
| $x_1$ | $-\frac{1}{3}$ | $\frac{2}{3}$ | 0 | $\frac{14}{3}$ |
| $x_6$ | 1 | $-1$ | 1 | 0 |

计算检验数得

$$(z_3-c_3,z_4-c_4,z_5-c_5)=\left(-1,-\frac{2}{3},-\frac{2}{3}\right).$$

由于所有的检验数 $z_j-c_j\leqslant 0$，故已经达到最优解. 最优解为

$$\boldsymbol{x}=(x_1,x_2,x_3)^{\mathrm{T}}=\left(\frac{14}{3},\frac{13}{3},0\right)^{\mathrm{T}}.$$

目标函数最优值　$f_{\min}=-\frac{26}{3}$.

## 1.4 有界变量单纯形法

### 1.4.1 基本概念

在许多实际问题中，线性规划的变量是有上、下界限制，一般形式为

$$\left.\begin{array}{ll}\min & \boldsymbol{cx};\\ \text{s.t.} & \boldsymbol{Ax}=\boldsymbol{b},\\ & \boldsymbol{L}\leqslant \boldsymbol{x}\leqslant \boldsymbol{U},\end{array}\right\}\tag{1.23}$$

**定义 1.4.1** 将上式中的 $\boldsymbol{A}$ 分解为 $(\boldsymbol{B}\ \boldsymbol{N}_1\ \boldsymbol{N}_2)$，其中，$\boldsymbol{B}$ 是 $m$ 阶满秩方阵. 相应地

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\ \boldsymbol{x}_{N_1}\\ \boldsymbol{x}_{N_2}\end{bmatrix},\quad \boldsymbol{L}=\begin{bmatrix}\boldsymbol{L}_B\\ \boldsymbol{L}_{N_1}\\ \boldsymbol{L}_{N_2}\end{bmatrix},\quad \boldsymbol{U}=\begin{bmatrix}\boldsymbol{U}_B\\ \boldsymbol{U}_{N_1}\\ \boldsymbol{U}_{N_2}\end{bmatrix}.$$

若 $\boldsymbol{x}$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解，且令 $\boldsymbol{x}_{N_1}=\boldsymbol{L}_{N_1}$，$\boldsymbol{x}_{N_2}=\boldsymbol{U}_{N_2}$，于是 $\boldsymbol{x}_B=\boldsymbol{B}^{-1}\boldsymbol{b}-\boldsymbol{B}^{-1}\boldsymbol{N}_1\boldsymbol{L}_{N_1}-\boldsymbol{B}^{-1}\boldsymbol{N}_2\boldsymbol{U}_{N_2}$. 若 $\boldsymbol{x}_B$ 满足 $\boldsymbol{L}_B\leqslant\boldsymbol{x}_B\leqslant\boldsymbol{U}_B$，则称

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_B\\ \boldsymbol{x}_{N_1}\\ \boldsymbol{x}_{N_2}\end{bmatrix}=\begin{bmatrix}\boldsymbol{B}^{-1}\boldsymbol{b}-\boldsymbol{B}^{-1}\boldsymbol{N}_1\boldsymbol{L}_{N_1}-\boldsymbol{B}^{-1}\boldsymbol{N}_2\boldsymbol{U}_{N_2}\\ \boldsymbol{L}_{N_1}\\ \boldsymbol{U}_{N_2}\end{bmatrix}$$

为(1.23)的一个基本可行解. 称 $\boldsymbol{B}$ 为基，称 $\boldsymbol{x}_B$ 为基变量，$\boldsymbol{x}_{N_1}$ 为**下非基变量**(lower nonbasic variables)，指标集为 $L$；$\boldsymbol{x}_{N_2}$ 为**上非基变量**(upper nonbasic variables)，指标集为 $U$.

**定义 1.4.2** 在 1.23 中，若 $\boldsymbol{L}_B<\boldsymbol{x}_B<\boldsymbol{U}_B$，则称 $\boldsymbol{x}$ 为**非退化的基本可行解**；否则，称 $\boldsymbol{x}$ 为**退化的基本可行解**.

**定理 1.4.3** 若(1.23)有有限的最优值，则一定在基本可行解处达到.

### 1.4.2 有界变量单纯形法原理

本节介绍一种不需扩大系数矩阵而求解(1.23)的单纯形法.

有界变量单纯形法的基本思想仍然是给定初始基本可行解后,从一个基本可行解出发,求改进的基本可行解,直至求出最优解或判定问题无界.

对于(1.23),将 $\boldsymbol{A}$ 分解为$(\boldsymbol{B}\ \boldsymbol{N}_1\ \boldsymbol{N}_2)$,其中,$\boldsymbol{B}$ 为基矩阵,$\boldsymbol{N}_1,\boldsymbol{N}_2$ 为非基矩阵,$\boldsymbol{N}_1$ 对应下非基变量,$\boldsymbol{N}_2$ 对应上非基变量.于是有

$$\begin{aligned}\boldsymbol{x}_B &= \boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{N}_1\boldsymbol{x}_{N_1} - \boldsymbol{B}^{-1}\boldsymbol{N}_2\boldsymbol{x}_{N_2} \\ &= \boldsymbol{B}^{-1}\boldsymbol{b} - \sum_{j\in L}\bar{\boldsymbol{a}}_j x_j - \sum_{j\in U}\bar{\boldsymbol{a}}_j x_j. \end{aligned} \tag{1.24}$$

$$\begin{aligned} f_0 &= \boldsymbol{C}_B\boldsymbol{x}_B + \boldsymbol{C}_{N_1}\boldsymbol{x}_{N_1} + \boldsymbol{C}_{N_2}\boldsymbol{x}_{N_2} \\ &= \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} - (\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_1 - \boldsymbol{C}_{N_1})\boldsymbol{x}_{N_1} - (\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_2 - \boldsymbol{C}_{N_2})\boldsymbol{x}_{N_2} \\ &= \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} - \sum_{j\in L}(z_j - c_j)x_j - \sum_{j\in U}(z_j - c_j)x_j. \end{aligned} \tag{1.25}$$

这里,$z_j-c_j$ 仍然称为检验数.由此得单纯形表如下:

**表 1.34　有界变量单纯形表**

| | $\boldsymbol{x}_B$ | $\boldsymbol{x}_{N_1}$ | $\boldsymbol{x}_{N_2}$ | |
|---|---|---|---|---|
| $\boldsymbol{x}_B$ | $\boldsymbol{I}$ | $\boldsymbol{B}^{-1}\boldsymbol{N}_1$ | $\boldsymbol{B}^{-1}\boldsymbol{N}_2$ | $\bar{\boldsymbol{b}}$ |
| $f$ | $\boldsymbol{0}$ | $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_1-\boldsymbol{C}_{N_1}$ | $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_2-\boldsymbol{C}_{N_2}$ | $f_0$ |

表中,$\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\boldsymbol{b}-\boldsymbol{B}^{-1}\boldsymbol{N}_1\boldsymbol{x}_{N_1}-\boldsymbol{B}^{-1}\boldsymbol{N}_2\boldsymbol{x}_{N_2}$.

在表 1.34 对应的基本可行解中,下非基变量的取值都达到了下界,不能再减少;上非基变量的取值都达到了上界,不能再增加.因此,由(1.25)可知,若有 $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_1-\boldsymbol{C}_{N_1}\leqslant\boldsymbol{0}$ 和 $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_2-\boldsymbol{C}_{N_2}\geqslant\boldsymbol{0}$,则目标函数值不能再减少.于是可得(1.23)的最优解判别准则如下:

**定理 1.4.4**　设 $\boldsymbol{x}$ 是(1.23)的一个基本可行解,对应的单纯

形表为表 1.34. 若有 $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_1-\boldsymbol{C}_{N_1}\leqslant\boldsymbol{0}$ 和 $\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{N}_2-\boldsymbol{C}_{N_2}\geqslant\boldsymbol{0}$，则 $\boldsymbol{x}$ 是 1.23 的最优解.

如果上述最优解判别准则不满足，需要求改进的基本可行解. 这时，确定进基变量和离基变量的基本原则如下：

(1) 设存在 $j\in L, z_j-c_j>0$. 于是可给 $x_j$ 增加值 $\delta$，使目标函数值减少. 令 $x_j=l_j+\delta$，其余非基变量取值不变化，则有

$$\boldsymbol{x}_B=\bar{\boldsymbol{b}}-\delta\bar{\boldsymbol{a}}_j,\quad f=f_0-\delta(z_j-c_j).$$

(2) 设存在 $j\in U, z_j-c_j<0$. 这时可使 $x_j$ 减少一个量 $\delta$. 令 $x_j=u_j-\delta$. 则有

$$\boldsymbol{x}_B=\bar{\boldsymbol{b}}+\delta\bar{\boldsymbol{a}}_j,\ f=f_0+\delta(z_j-c_j).$$

综合(1)，(2)，令

$$\lambda=\max\{\max_{j\in L}(z_j-c_j),\max_{j\in U}-(z_j-c_j)\},\qquad(1.26)$$

如果 $\lambda>0$，设该最大值在指标 $k$ 上达到，则 $x_k$ 可能为进基变量.

确定离基变量仍按最小比原则. 由于 $\delta\leqslant u_k-l_k, \boldsymbol{L}_B\leqslant\boldsymbol{x}_B=\bar{\boldsymbol{b}}-\delta\bar{\boldsymbol{a}}_k\leqslant\boldsymbol{U}_B$，故

若 $k\in L$，则

$$\delta\leqslant\delta_1=\min\left\{\frac{\bar{b}_i-l_{B_i}}{\bar{a}_{ik}}\middle|\bar{a}_{ik}>0\right\}=\frac{\bar{b}_r-l_{B_r}}{\bar{a}_{rk}}.\qquad(1.27)$$

$$\delta\leqslant\delta_2=\min\left\{\frac{U_{B_i}-\bar{b}_i}{(-\bar{a}_{ik})}\middle|\bar{a}_{ik}<0\right\}=\frac{U_{B_r}-\bar{b}_r}{(-\bar{a}_{rk})}.\qquad(1.28)$$

若 $k\in U$，则

$$\delta\leqslant\delta_1=\min\left\{\frac{\bar{b}_i-l_{B_i}}{(-\bar{a}_{ik})}\middle|\bar{a}_{ik}<0\right\}=\frac{\bar{b}_r-l_{B_r}}{(-\bar{a}_{rk})}.\qquad(1.29)$$

$$\delta\leqslant\delta_2=\min\left\{\frac{U_{B_i}-\bar{b}_i}{\bar{a}_{ik}}\middle|\bar{a}_{ik}>0\right\}=\frac{U_{B_r}-\bar{b}_r}{\bar{a}_{rk}}.\qquad(1.30)$$

取 $\delta=\min\{u_k-l_k,\delta_1,\delta_2\}$，当 $\delta\neq+\infty$ 时，若 $\delta=\delta_1$ 或 $\delta_2$，则 $x_{B_r}$ 为离基变量；否则，$x_k$ 仍为非基变量，只是取值有变化.

### 1.4.3 有界变量单纯形法的计算步骤

(1) 给定初始基本可行解,构造单纯形表 1.34,得指标集 $L$ 和 $U$.

(2) 按照(1.26)求 $\lambda$. 若 $\lambda\leqslant 0$,停止计算,现行基本可行解为最优解;否则,设该最大值在指标 $k$ 上达到. 若 $k\in L$,转(3);若 $k\in U$,转(4).

(3) 按照(1.27)和(1.28)计算 $\delta_1$ 和 $\delta_2$,并求 $\delta=\min\{u_k-l_k,\delta_1,\delta_2\}$. 若 $\delta=+\infty$,停止计算,目标函数值无界;否则,若 $\delta\neq u_k-l_k$, $x_k$ 进基, $x_{B_r}$ 离基. 以 $\bar{a}_{rk}$ 为主元,对左端进行消元法运算. 并且,按下面公式修改右端:

$$x_k=l_k+\delta. \tag{1.31}$$

$$\boldsymbol{x}_B=\bar{\boldsymbol{b}}-\delta\bar{\boldsymbol{a}}_k. \tag{1.32}$$

$$f=f_0-\delta(z_k-c_k), \tag{1.33}$$

转(5);若 $\delta=u_k-l_k$,令 $x_k=u_k$,修改右端,单纯形表上其余数据不变化,转(5).

(4) 按照(1.29)和(1.30)计算 $\delta_1$ 和 $\delta_2$,并求 $\delta=\min\{u_k-l_k,\delta_1,\delta_2\}$. 若 $\delta=+\infty$,停止计算,目标函数值无界;否则,若 $\delta\neq u_k-l_k$, $x_k$ 进基, $x_{B_r}$ 离基. 以 $\bar{a}_{rk}$ 为主元,对左端进行消元法运算. 按下面公式修改右端:

$$x_k=u_k-\delta. \tag{1.34}$$

$$\boldsymbol{x}_B=\bar{\boldsymbol{b}}+\delta\bar{\boldsymbol{a}}_k. \tag{1.35}$$

$$f=f_0+\delta(z_k-c_k), \tag{1.36}$$

转(6);若 $\delta=u_k-l_k$,令 $x_k=l_k$,修改右端,单纯形表上其余数据不变化,转(6).

(5) 修改指标集, $L\setminus\{k\}\to L$, $U\cup\{k\}\to U$,返回(2).

(6) 修改指标集, $L\cup\{k\}\to L$, $U\setminus\{k\}\to U$,返回(2).

**例 1.4.5** 求解下列有界变量问题

$$
\begin{aligned}
\min \quad & -2x_1 - 4x_2 - x_3 \\
\text{s.t.} \quad & 2x_1 + x_2 + x_3 + x_4 = 10, \\
& x_1 + x_2 - x_3 + x_5 = 4, \\
& 0 \leqslant x_1 \leqslant 4, \\
& 0 \leqslant x_2 \leqslant 6, \\
& 1 \leqslant x_3 \leqslant 4, \\
& x_4 \geqslant 0,\ x_5 \geqslant 0.
\end{aligned}
$$

**解** 初始基变量为 $x_4$ 和 $x_5$，下非基变量为 $x_1$，$x_2$ 和 $x_3$. 初始基本可行解为

$$
\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_B \\ \\ \boldsymbol{x}_{N_l} \end{bmatrix} = \begin{bmatrix} 9 \\ 5 \\ 0 \\ 0 \\ 1 \end{bmatrix},
$$

相应的目标函数值 $f_0 = -1$. 下标集 $L = \{1,2,3\}$. 构造初始单纯形表(**注意**，用在变量上方标记"$l$"或"$u$"表示其为下非基变量或上非基变量).

**表 1.35 迭代 1**

| | $l$ | $l$ | $l$ | | | |
|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
| $x_4$ | 2 | 1 | 1 | 1 | 0 | 9 |
| $x_5$ | 1 | $\boxed{1}$ | $-1$ | 0 | 1 | 5 |
| $f$ | 2 | 4 | 1 | 0 | 0 | $-1$ |

由于 $\max\limits_{j \in L} \{z_j - c_j\} = 4 = z_2 - c_2$，故 $k = 2$，即可以让 $x_2$ 的值增加.

按照(1.27)计算 $\delta_1$：

$$\delta_1 = \min\left\{\frac{9-0}{1},\frac{5-0}{1}\right\} = 5 = \frac{\bar{b}_2 - l_{B_2}}{\bar{a}_{22}},$$

相应地，$x_{B_2}=x_5$，即 $r=2$. 又知 $\delta_2=+\infty$，$u_2-l_2=6$. 故有

$$\delta = \min\{u_2 - l_2,\delta_1,\delta_2\} = \min\{6,5,+\infty\} = 5.$$

因为 $\delta \neq u_2-l_2$，所以，$x_2$ 进基，$x_5$ 离基. 按(1.31)～(1.33)修改右端，计算结果为

$$x_2 = 0 + 5 = 5,$$

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 9 \\ 5 \end{bmatrix} - 5\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \end{bmatrix},$$

$$f = -1 - 5 \times 4 = -21.$$

以 $\bar{a}_{22}$ 为主元，对左端进行消元法运算得表 1.36.

**表 1.36　迭代 2**

| | $l$<br>$x_1$ | $x_2$ | $l$<br>$x_3$ | $x_4$ | $l$<br>$x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_4$ | 1 | 0 | 2 | 1 | $-1$ | 4 |
| $x_2$ | 1 | 1 | [$-1$] | 0 | 1 | 5 |
| $f$ | $-2$ | 0 | 5 | 0 | $-4$ | $-21$ |

所有的非基变量均为下非基变量. 这时，$L=\{1,3,5\}$，$U=\varnothing$.

由于$\max\limits_{j\in L}\{z_j-c_j\}=5=z_3-c_3$，因此，$k=3$，故可让 $x_3$ 的值增加.

按照(1.27)和(1.28)计算 $\delta_1$ 和 $\delta_2$：

$$\delta_1 = \frac{4-0}{2} = 2, \text{对应于 } x_{B_r} = x_{B_1} = x_4, r = 1.$$

$$\delta_2 = \frac{6-5}{1} = 1, \text{对应于 } x_{B_r} = x_{B_2} = x_2, r = 2.$$

又知 $u_3-l_3=3$，故有

$\delta=\min\{3,2,1\}=1=\delta_2$. 所以，$x_3$ 进基，$x_2$ 离基. 按(1.31)～(1.33)修改右端，计算结果为

$$x_3 = 1 + 1 = 2,$$

$$\begin{bmatrix} x_4 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \end{bmatrix},$$

$$f = -21 - 1 \times 5 = -26.$$

以 $\bar{a}_{23}=-1$ 为主元，对左端进行消元法运算得表 1.37.

**表 1.37　迭代 3**

| | $l$<br>$x_1$ | $u$<br>$x_2$ | $x_3$ | $x_4$ | $l$<br>$x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_4$ | [3] | 2 | 0 | 1 | 1 | 2 |
| $x_3$ | $-1$ | $-1$ | 1 | 0 | $-1$ | 2 |
| $f$ | 3 | 5 | 0 | 0 | 1 | $-26$ |

**注意**，$x_2$ 为上非基变量. 即 $L=\{1,5\}$，$U=\{2\}$.

由于 $\max\{\max\limits_{j\in L}(z_j-c_j),\max\limits_{j\in U}-(z_j-c_j)\}=3=z_1-c_1$，故 $k=1$，因而可以让 $x_1$ 的值增加.

按照(1.27)和(1.28)计算 $\delta_1$ 和 $\delta_2$：

$$\delta_1 = \frac{2-0}{3} = \frac{2}{3}, \text{对应于 } x_{B_r} = x_{B_1} = x_4, r = 1.$$

$$\delta_2 = \frac{4-2}{1} = 2, \text{对应于 } x_{B_r} = x_{B_2} = x_3, r = 2.$$

又 $u_1-l_1=4$，故有

$\delta=\min\left\{4,\frac{2}{3},2\right\}=\frac{2}{3}=\delta_1$. 所以，$x_1$ 进基，$x_4$ 离基. 按(1.31)～(1.33)修改右端，计算结果为

$$x_1 = 0 + \frac{2}{3} = \frac{2}{3},$$

$$\begin{bmatrix} x_4 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{8}{3} \end{bmatrix}.$$

$$f = -26 - \frac{2}{3} \times 3 = -28.$$

以 $\bar{a}_{11} = 3$ 为主元,对左端进行消元法运算得表 1.38.

**表 1.38 迭代 4**

| | $x_1$ | $x_2$ (u) | $x_3$ | $x_4$ (l) | $x_5$ (l) | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_1$ | 1 | $\frac{2}{3}$ | 0 | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| $x_3$ | 0 | $-\frac{1}{3}$ | 1 | $\frac{1}{3}$ | $-\frac{2}{3}$ | $\frac{8}{3}$ |
| $f$ | 0 | 3 | 0 | $-1$ | 0 | $-28$ |

因为对于上非基变量有 $z_j - c_j \geqslant 0$,而对于下非基变量有 $z_j - c_j \leqslant 0$,故已得到最优解 $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)^{\mathrm{T}} = \left(\frac{2}{3}, 6, \frac{8}{3}, 0, 0\right)^{\mathrm{T}}$,目标函数最优值 $f = -28$.

# 1.5 最优性条件和对偶问题

## 1.5.1 Kuhn-Tucker 条件

本节给出线性规划问题的 Kuhn-Tucker 条件,简称 K-T 条件. 这些条件对于构造线性规划对偶问题及其解法是十分有用的.

**1 不等式约束问题的 K-T 条件**

设有线性规划问题

$$\left.\begin{aligned} \min \quad & \boldsymbol{cx}; \\ \text{s.t.} \quad & \boldsymbol{Ax} \geqslant \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned}\right\} \tag{1.37}$$

则 $\boldsymbol{x}^*$ 是该问题的最优解的充要条件是，存在 $n$ 维向量 $\boldsymbol{v}$ 和 $m$ 维向量 $\boldsymbol{w}$，使下列 K-T 条件成立：

$$\boldsymbol{A}\boldsymbol{x}^* \geqslant \boldsymbol{b},\quad \boldsymbol{x}^* \geqslant \boldsymbol{0}. \tag{1.38}$$

$$\boldsymbol{c} - \boldsymbol{w}\boldsymbol{A} - \boldsymbol{v} = \boldsymbol{0}, \boldsymbol{w} \geqslant \boldsymbol{0}, \boldsymbol{v} \geqslant \boldsymbol{0}. \tag{1.39}$$

$$\boldsymbol{w}(\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}) = 0, \boldsymbol{v}\boldsymbol{x}^* = 0. \tag{1.40}$$

**2 等式约束问题的 K-T 条件**

设有线性规划问题

$$\begin{aligned} \min \quad & \boldsymbol{c}\boldsymbol{x}; \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned} \tag{1.41}$$

则 $\boldsymbol{x}^*$ 是该问题的最优解的充要条件是，存在 $n$ 维向量 $\boldsymbol{v}$ 和 $m$ 维向量 $\boldsymbol{w}$，满足 K-T 条件：

$$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b},\quad \boldsymbol{x}^* \geqslant \boldsymbol{0}. \tag{1.42}$$

$$\boldsymbol{c} - \boldsymbol{w}\boldsymbol{A} - \boldsymbol{v} = \boldsymbol{0}, \boldsymbol{v} \geqslant 0. \tag{1.43}$$

$$\boldsymbol{v}\boldsymbol{x}^* = 0. \tag{1.44}$$

这就是标准形式线性规划问题的 K-T 条件.

**注意**，它与不等式约束问题 K-T 条件的主要区别是，这里对 $\boldsymbol{w}$ 没有非负限制的要求.

标准形式线性规划问题的 K-T 条件也可写成如下的形式：

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\quad \boldsymbol{x} \geqslant 0. \tag{1.45}$$

$$\boldsymbol{w}\boldsymbol{A} \leqslant \boldsymbol{c}. \tag{1.46}$$

$$(\boldsymbol{w}\boldsymbol{A} - \boldsymbol{c})\boldsymbol{x} = 0. \tag{1.47}$$

### 1.5.2 对偶问题的表示

**1 对称的对偶规划**

给定线性规划问题

$$\begin{aligned} \min \quad & \boldsymbol{c}\boldsymbol{x}; \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned} \tag{1.48}$$

将(1.48)的 K-T 条件中的(1.39)改写为

$$\boldsymbol{wA} \leqslant \boldsymbol{c},\ \boldsymbol{w} \geqslant \boldsymbol{0},$$

作为约束条件,把 $\boldsymbol{w}$ 作为变量,构造线性规划问题

$$\begin{aligned} \max \quad & \boldsymbol{wb}; \\ \text{s.t.} \quad & \boldsymbol{wA} \leqslant \boldsymbol{c}, \\ & \boldsymbol{w} \geqslant \boldsymbol{0}. \end{aligned} \tag{1.49}$$

称(1.48)为**原问题**(primal problem),记成(P). 称(1.49)为(1.48)的**对偶问题**(dual problem),记成(D). (1.48)和(1.49)称为对称的对偶规划.

### 2 非对称的对偶规划

原问题为

$$\begin{aligned} \min \quad & \boldsymbol{cx}; \\ \text{s.t.} \quad & \boldsymbol{Ax} = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned} \tag{1.50}$$

其对偶问题为

$$\begin{aligned} \max \quad & \boldsymbol{wb}; \\ \text{s.t.} \quad & \boldsymbol{wA} \leqslant \boldsymbol{c}, \\ & \boldsymbol{w}\ \text{无限制}. \end{aligned} \tag{1.51}$$

### 3 混合型对偶规划

原问题为

$$\begin{aligned} \min \quad & \boldsymbol{cx}; \\ \text{s.t.} \quad & \boldsymbol{A}_1\boldsymbol{x} \geqslant \boldsymbol{b}_1, \\ & \boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{b}_2, \\ & \boldsymbol{A}_3\boldsymbol{x} \leqslant \boldsymbol{b}_3, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned} \tag{1.52}$$

对偶问题为

$$
\begin{aligned}
\max \quad & \boldsymbol{w}_1\boldsymbol{b}_1 + \boldsymbol{w}_2\boldsymbol{b}_2 + \boldsymbol{w}_3\boldsymbol{b}_3 ; \\
\text{s.t.} \quad & \boldsymbol{w}_1\boldsymbol{A}_1 + \boldsymbol{w}_2\boldsymbol{A}_2 + \boldsymbol{w}_3\boldsymbol{A}_3 \leqslant \boldsymbol{c}, \\
& \boldsymbol{w}_1 \geqslant \boldsymbol{0}, \\
& \boldsymbol{w}_3 \leqslant \boldsymbol{0}, \\
& \boldsymbol{w}_2 \ \text{无限制}.
\end{aligned}
\tag{1.53}
$$

综上所述,原问题与对偶问题之间的关系如表 1.39:

**表 1.39**

| 原问题(P)min | | 对偶问题(D)max | |
|---|---|---|---|
| 变 | $\geqslant 0$ | 行 | $\leqslant$ |
| | $\leqslant 0$ | 约 | $\geqslant$ |
| 量 | 无限制 | 束 | $=$ |
| 行 | $\geqslant$ | 变 | $\geqslant 0$ |
| 约 | $\leqslant$ | | $\leqslant 0$ |
| 束 | $=$ | 量 | 无限制 |

给定线性规划,按照表 1.39 的对应关系便可以写出相应的对偶规划.

**例 1.5.1** 写出下列线性规划的对偶规划

$$
\begin{aligned}
(\mathrm{P}) \quad \min \quad & 6x_1 + 8x_2 ; \\
\text{s.t.} \quad & 3x_1 + x_2 \geqslant 4, \\
& 5x_1 + 2x_2 \geqslant 7, \\
& x_1 \geqslant 0,\ x_2 \geqslant 0.
\end{aligned}
$$

**解** 对偶问题为

$$
\begin{aligned}
(\mathrm{D}) \quad \max \quad & 4w_1 + 7w_2 ; \\
\text{s.t.} \quad & 3w_1 + 5w_2 \leqslant 6, \\
& w_1 + 2w_2 \leqslant 8, \\
& w_1 \geqslant 0,\ w_2 \geqslant 0.
\end{aligned}
$$

**例 1.5.2** 写出下列线性规划的对偶问题

$$\begin{aligned}
\min \quad & -4x_1 - 5x_2 - 7x_3 + x_4 ; \\
\text{s.t.} \quad & x_1 + x_2 + 2x_3 - x_4 \geqslant 1, \\
& 2x_1 - 6x_2 + 3x_3 + x_4 \leqslant -3, \\
& x_1 + 4x_2 + 3x_3 + 2x_4 = -5, \\
& x_1, x_2, x_4 \geqslant 0.
\end{aligned}$$

**解** 对偶问题为

$$\begin{aligned}
\max \quad & w_1 - 3w_2 - 5w_3 ; \\
\text{s.t.} \quad & w_1 + 2w_2 + w_3 \leqslant -4, \\
& w_1 - 6w_2 + 4w_3 \leqslant -5, \\
& 2w_1 + 3w_2 + 3w_3 = -7, \\
& -w_1 + w_2 + 2w_3 \leqslant 1, \\
& w_1 \geqslant 0, \\
& w_2 \leqslant 0, \\
& w_3 \text{ 无限制}.
\end{aligned}$$

**定理 1.5.3** 对偶问题的对偶问题就是原问题.

### 1.5.3 对偶原理

本节叙述一对对偶线性规划问题之间的关系.由于各种形式的对偶规划都可以转化为对称的对偶规划,所以仅研究(1.48)和(1.49).

**定理 1.5.4** 设 $\boldsymbol{x}$ 和 $\boldsymbol{w}$ 分别为(1.48)和(1.49)的可行解,则有 $\boldsymbol{cx} \geqslant \boldsymbol{wb}$.

**定理 1.5.5** (1.48)和(1.49)同时有最优解的充要条件是它们同时有可行解.并且,若其中一个问题无界,则另一个问题无解.

**定理 1.5.6** 设 $\boldsymbol{x}^*$ 和 $\boldsymbol{w}^*$ 分别为(1.48)和(1.49)的可行解,则它们分别为(1.48)和(1.49)的最优解的充要条件是 $\boldsymbol{cx}^* = \boldsymbol{w}^*\boldsymbol{b}$.

**定理 1.5.7** 设(1.48)有最优解 $\boldsymbol{x}^*$,则(1.49)有最优解 $\boldsymbol{w}^*$,

且 $\boldsymbol{c}\boldsymbol{x}^*=\boldsymbol{w}^*\boldsymbol{b}$.

**定理 1.5.8** 设 $\boldsymbol{x}^*$ 和 $\boldsymbol{w}^*$ 分别为(1.48)和(1.49)的最优解,则有

$$\boldsymbol{w}^*(A\boldsymbol{x}^*-\boldsymbol{b})=0, \tag{1.54}$$

$$(\boldsymbol{c}-\boldsymbol{w}^*A)\boldsymbol{x}^*=0. \tag{1.55}$$

或

$$w_i^*(\boldsymbol{a}^{(i)}\boldsymbol{x}^*-b_i)=0,\ i=1,\cdots,m,$$

$$(c_j-\boldsymbol{w}^*\boldsymbol{a}_j)x_j^*=0,j=1,\cdots,n.$$

定理 1.5.8 表明,若两对偶规划的最优解满足:(1.54)(1.55).则称,两对偶规划的最优解满足**互补松弛条件**(complementary slackness condition).

**定理 1.5.9** 若(1.48)有非退化的最优基本可行解,则(1.49)有唯一的最优解.

综上所述,对于两对偶规划,得知其中一个问题的最优解时,可以利用互补松弛条件求出另一个问题的最优解.例如,当(1.48)只有两个约束时,它的对偶问题(1.49)只有两个变量,可用图解法求得其最优解 $\boldsymbol{w}^*$ 和目标函数最优值 $\boldsymbol{w}^*\boldsymbol{b}$.然后,再利用互补松弛条件很容易求得原问题的最优解 $\boldsymbol{x}^*$,并且(P)的最优值也等于 $\boldsymbol{w}^*\boldsymbol{b}$.如果(D)无界,则(P)无解.

**例 1.5.10** 求解下列线性规划

$$\begin{aligned}
\text{(P)}\quad \min\ & 2x_1+3x_2+5x_3+2x_4+3x_5;\\
\text{s.t.}\ & x_1+x_2+2x_3+x_4+3x_5\geqslant 4,\\
& 2x_1-2x_2+3x_3+x_4+x_5\geqslant 3,\\
& x_j\geqslant 0,\ j=1,2,\cdots,5.
\end{aligned}$$

**解** 这问题的对偶问题是

$$
\begin{aligned}
(\mathrm{D})\quad & \max\ 4w_1+3w_2\ ;\\
\text{s.t.}\quad & w_1+2w_2\leqslant 2,\\
& w_1-2w_2\leqslant 3,\\
& 2w_1+3w_2\leqslant 5,\\
& w_1+w_2\leqslant 2,\\
& 3w_1+w_2\leqslant 3,\\
& w_1\geqslant 0,\ w_2\geqslant 0.
\end{aligned}
$$

利用图解法求解(D).

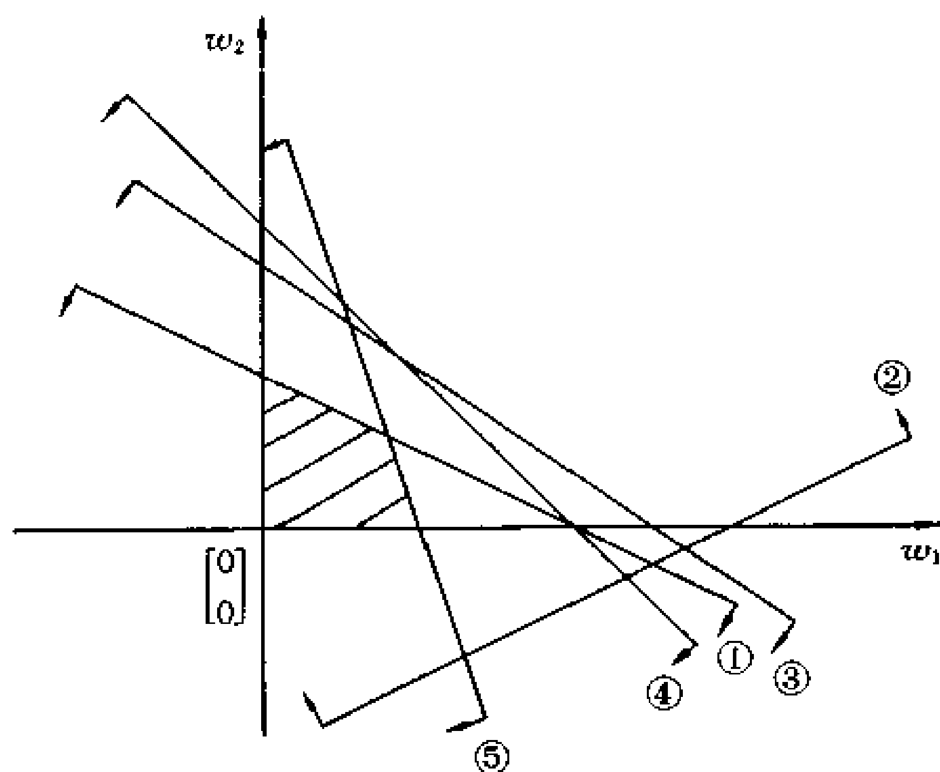



图 1.2

得最优解为 $\boldsymbol{w}^*=\left(\dfrac{4}{5},\dfrac{3}{5}\right)^{\mathrm{T}}$，目标函数最优值为 5.

由于对 $\boldsymbol{w}^*$ 来说，(D)的第 2,3,4 个行约束是非临界的，故若令(P)的最优解为 $\boldsymbol{x}^*=(x_1^*,x_2^*,x_3^*,x_4^*,x_5^*)^{\mathrm{T}}$，则根据互补松弛条

件便有 $x_2^*=x_3^*=x_4^*=0$. 又由于 $w_1^*, w_2^*>0$，所以(P)的行约束成立等式

$$\begin{cases} x_1^* + 3x_5^* = 4, \\ 2x_1^* + x_5^* = 3. \end{cases}$$

解此方程组得 $x_1^*=x_5^*=1$. 于是，得到(P)的最优解为 $\boldsymbol{x}^*=(1,0,0,0,1)^{\mathrm{T}}$，目标函数最优值是 5.

### 1.5.4 对偶单纯形法

首先介绍两个概念. 考虑标准形式线性规划及其对偶规划，即(1.50)和(1.51).

**定义 1.5.11** 设 $\boldsymbol{A}=(\boldsymbol{B}\ \boldsymbol{N})$，其中，$\boldsymbol{B}$ 是满秩方阵. 相应地，$\boldsymbol{c}=(\boldsymbol{c}_B,\boldsymbol{c}_N)$，则 $\boldsymbol{w}\boldsymbol{B}=\boldsymbol{c}_B$ 的解 $\boldsymbol{w}=\boldsymbol{c}_B\boldsymbol{B}^{-1}$ 称为对偶问题的一个**基本解**，$\boldsymbol{B}$ 称为与基本解对应的**基**；如果 $\boldsymbol{w}$ 还满足 $\boldsymbol{w}\boldsymbol{N}\leqslant\boldsymbol{c}_N$，则称 $\boldsymbol{w}$ 为对偶问题的**基本可行解**.

**定义 1.5.12** 设 $\boldsymbol{B}$ 是(P)的基，对应的基本解为 $\boldsymbol{x}$，若对应的检验数全部非正，即 $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{c}\leqslant\boldsymbol{0}$，则称 $\boldsymbol{x}$ 为(P)的一个**正则解**(regular solution).

由定义 1.5.11 与定义 1.5.12 可知，原问题的正则解 $\boldsymbol{x}$ 与对偶问题的基本可行解 $\boldsymbol{w}$ 是一一对应的，它们由同一个基 $\boldsymbol{B}$ 所决定.

**对偶单纯形法**(dual simplex method)的基本思想是：从对偶问题的一个基本可行解迭代到另一个基本可行解，且使目标函数值递增. 也就是说，从原问题的一个正则解迭代到另一个正则解，当迭代到 $\boldsymbol{B}^{-1}\boldsymbol{b}\geqslant\boldsymbol{0}$，即正则解满足原问题的可行性时，即原问题与对偶问题同时求得可行解，也就同时求得最优解.

对偶单纯形法首先确定离基变量：一般，取最小的那个 $\bar{b}_r<0$ 对应的 $x_{B_r}$ 为离基变量. 然后确定进基变量. 确定进基变量的原则是使迭代后得到的解仍然是原问题的正则解.

设 $x_k$ 为进基变量. 因为迭代后的目标函数值为

$$f=\bar{f}-\sum_{\substack{j\in N_B\\ j\neq k}}\left[(z_j-c_j)-\frac{z_k-c_k}{\bar{a}_{rk}}\bar{a}_{rj}\right]x_j$$

$$-\frac{z_k-c_k}{\bar{a}_{rk}}\bar{b}_r+\frac{z_k-c_k}{\bar{a}_{rk}}x_{B_r}$$

所以,为了使迭代后的解仍为正则解,并且使目标函数值增大,确定 $x_k$ 的指标 $k$ 必须满足下列两个条件:

$$\begin{cases}-\dfrac{z_k-c_k}{\bar{a}_{rk}}\leqslant 0, & (1.56)\\[2ex] (z_j-c_j)-\dfrac{z_k-c_k}{\bar{a}_{rk}}\bar{a}_{rj}\leqslant 0 \quad (j\in N_B,j\neq k). & (1.57)\end{cases}$$

(1.56)要求 $\bar{a}_{rk}<0$. 若 $\bar{a}_{rj}\geqslant 0$,则(1.57)一定成立;若 $\bar{a}_{rj}<0$,则(1.57)要求有

$$\frac{z_k-c_k}{\bar{a}_{rk}}\leqslant\frac{z_j-c_j}{\bar{a}_{rj}}\quad (j\in N_B,j\neq k).$$

故进基变量 $x_k$ 可由下式确定:

$$\frac{z_k-c_k}{\bar{a}_{rk}}=\min\left\{\frac{z_j-c_j}{\bar{a}_{rj}}\,\middle|\,\bar{a}_{rj}<0\right\}.$$

对偶单纯形法的迭代过程也可以在表上进行. 对偶单纯形表和原始单纯形表形式上完全一样. 二者区别在于,对偶单纯形法是保证全部检验数 $z_j-c_j\leqslant 0$,但不保证 $\bar{\boldsymbol{b}}\geqslant\boldsymbol{0}$.

对偶单纯形法的计算步骤:

(1) 找出一个初始对偶基本可行解,即初始正则解 $\boldsymbol{x}^{(0)}$,相应的基为 $\boldsymbol{B}$.

(2) 若 $\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\boldsymbol{b}\geqslant\boldsymbol{0}$,停止计算,已求得原问题的最优解;否则,计算

$$\bar{b}_r=\min\{\bar{b}_i\,|\,i=1,2,\cdots,m\}.$$

(3) 若 $\bar{a}_{rj}\geqslant 0$, $j=1,\cdots,n$,停止计算,原问题无解;否则,计算

$$\frac{z_k - c_k}{\bar{a}_{rk}} = \min\left\{\frac{z_j - c_j}{\bar{a}_{rj}} \middle| \bar{a}_{rj} < 0\right\} \tag{1.58}$$

(4) 以$\bar{a}_{rk}$为主元，进行消元法运算. 返回(2).

**例 1.5.13** 用对偶单纯形法解下列问题

$$\begin{aligned} \min \quad & 4x_1 + 6x_2 + 18x_3 ; \\ \text{s.t.} \quad & x_1 \qquad + 3x_3 \geqslant 3, \\ & \qquad x_2 + 2x_3 \geqslant 5, \\ & x_1, x_2, x_3 \geqslant 0. \end{aligned}$$

**解** 引入松弛变量 $x_4, x_5 \geqslant 0$，再将所得到的等式约束两端同乘以(−1)，就得到了原问题的一个正则解. 对应的初始单纯形表 1.40 如下：

**表 1.40**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_4$ | −1 | 0 | −3 | 1 | 0 | −3 |
| $x_5$ | 0 | [−1] | −2 | 0 | 1 | −5 |
| $f$ | −4 | −6 | −18 | 0 | 0 | 0 |

由于 $\bar{b}_2 = -5 < -3 = \bar{b}_1$，故 $x_{B_2} = x_5$ 为离基变量. 按照(1.58)求最小比，确定 $x_2$ 为进基变量. 以$\bar{a}_{22} = -1$ 为主元，进行消元法运算得表 1.41.

**表 1.41**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_4$ | −1 | 0 | −3 | 1 | 0 | −3 |
| $x_2$ | 0 | 1 | 2 | 0 | −1 | 5 |
| $f$ | −4 | 0 | −6 | 0 | −6 | 30 |

$\bar{b}_1 = -3 < 0$，$x_{B_1} = x_4$ 为离基变量. 由最小比，确定 $x_3$ 为进基变量. 以$\bar{a}_{13} = -3$ 为主元，进行消元法运算，得表 1.42.

表 1.42

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_3$ | $\frac{1}{3}$ | 0 | 1 | $-\frac{1}{3}$ | 0 | 1 |
| $x_2$ | $-\frac{2}{3}$ | 1 | 0 | $\frac{2}{3}$ | $-1$ | 3 |
| $f$ | $-2$ | 0 | 0 | $-2$ | $-6$ | 36 |

由于表 1.42 中 $\bar{\boldsymbol{b}}=(1,3)^{\mathrm{T}}>\boldsymbol{0}$，故已得原问题的最优解 $\boldsymbol{x}^*=(0,3,1)^{\mathrm{T}}$，目标函数最优值为 36.

### 1.5.5 原始-对偶单纯形法

原始-对偶单纯形法(primal-dual simplex method)同时对对偶问题和原问题第一阶段辅助问题求解. 它从对偶问题的一个可行解开始迭代，保持对偶可行性和互补松弛性，最终达到原问题的可行性，从而求得原问题的最优解.

考虑标准形式线性规划及其对偶规划，即(1.63)和(1.64). 假设已知(D)的一个可行解 $\boldsymbol{w}$. 于是，对一切 $j$ 有 $\boldsymbol{w}\boldsymbol{a}_j\leqslant c_j$. 为了求(P)的满足互补松弛性的可行解，设集合

$$Q=\{j\,|\,\boldsymbol{w}\boldsymbol{a}_j-c_j=0\},$$

即 $Q$ 为允许取正值的那些变量 $x_j$ 的下标集. 当 $j\notin Q$ 时，令 $x_j=0$；当 $j\in Q$ 时，对这些 $x_j$ 求解(P)的如下形式的辅助问题：

$$\begin{aligned}
\min\quad & g=\sum_{j\in Q}0x_j+\boldsymbol{e}\boldsymbol{x}_a;\\
\text{s.t.}\quad & \sum_{j\in Q}\boldsymbol{a}_jx_j+\boldsymbol{x}_a=\boldsymbol{b},\\
& x_j\geqslant 0\quad j\in Q,\\
& \boldsymbol{x}_a\geqslant\boldsymbol{0}.
\end{aligned}\tag{1.59}$$

(1.59)称为相应于对偶可行解 $\boldsymbol{w}$ 的**限定原问题**(restricted primal

problem).

设 $g_0$ 为(1.59)的最优目标值.若 $g_0=0$,说明已经求得(P)的可行解.又因为这个解满足互补松弛性条件,所以它就是(P)的最优解;若 $g_0>0$,说明尚未求得(P)的可行解,因而需再找(D)的另一个可行解 $\hat{w}$:要求 $\hat{w}$ 能使(1.59)的目标值减小,并且满足对偶可行性.为此求解(1.59)的对偶问题

$$
\begin{aligned}
\max \quad & \boldsymbol{vb}; \\
\text{s.t.} \quad & \boldsymbol{va}_j \leqslant 0 \quad j \in Q, \\
& \boldsymbol{v} \leqslant \boldsymbol{e}^{\mathrm{T}}.
\end{aligned} \tag{1.60}
$$

设 $\boldsymbol{v}^*$ 是(1.73)的最优解,则新的对偶可行解为

$$
\hat{\boldsymbol{w}} = \boldsymbol{w} + \theta \boldsymbol{v}^*. \tag{1.61}
$$

其中

$$
\theta = \min_{1\leqslant j\leqslant n}\left\{-\frac{\boldsymbol{wa}_j - c_j}{\boldsymbol{v}^*\boldsymbol{a}_j}\,\middle|\,\boldsymbol{v}^*\boldsymbol{a}_j > 0\right\}. \tag{1.62}
$$

求出新的对偶可行解 $\hat{\boldsymbol{w}}$ 后,修改 $Q$,再求解(1.72),如此迭代下去.如果限定原问题都不出现退化现象,则经过有限次迭代必能求得(P)的最优解或判定(P)无解.

原始-对偶单纯形法的计算步骤:

(1) 将线性规划变换为标准形式(1.50),求出其对偶规划(1.51)的一个可行解 $\boldsymbol{w}$.

(2) 令 $Q=\{j\mid \boldsymbol{wa}_j-c_j=0\}$,求解限定原问题(1.59).若其最优目标值 $g_0=0$,停止计算,已求得原问题的最优解;否则,转(3).

(3) 求解限定原问题的对偶问题(1.60),设其最优解为 $\boldsymbol{v}$.若 $\boldsymbol{vA}\leqslant\boldsymbol{0}$,停止计算,原问题无解;否则,按照(1.61)和(1.62)求出新的对偶可行解 $\hat{\boldsymbol{w}}$.以 $\hat{\boldsymbol{w}}$ 代替 $\boldsymbol{w}$,返回(2).

**例 1.5.14** 用原始-对偶单纯形法求解下列问题

$$
\begin{aligned}
\min \quad & 3x_1 + 2x_2 + x_3 + 2x_4 + 2x_5 \ ; \\
\text{s.t.} \quad & x_1 - x_2 + 2x_3 - x_4 + x_5 + 2x_6 = 1, \\
& -x_1 + 2x_2 + x_3 - 2x_4 - x_5 + x_6 = 3, \\
& 2x_1 + x_2 - x_3 + x_4 - 2x_5 + x_6 = 2, \\
& x_j \geqslant 0 \quad j = 1,2,\cdots,6.
\end{aligned}
$$

**解** 它的对偶规划是

$$
\begin{aligned}
\max \quad & w_1 + 3w_2 + 2w_3 \ ; \\
\text{s.t.} \quad & w_1 - w_2 + 2w_3 \leqslant 3, \\
& -w_1 + 2w_2 + w_3 \leqslant 2, \\
& 2w_1 + w_2 - w_3 \leqslant 1, \\
& -w_1 - 2w_2 + w_3 \leqslant 2, \\
& w_1 - w_2 - 2w_3 \leqslant 2, \\
& 2w_1 + w_2 + w_3 \leqslant 0.
\end{aligned}
$$

$\boldsymbol{w}=(0,0,0)$是它的一个初始可行解.这时,最后一个约束等式成立,故有 $Q=\{6\}$,对应的限定原问题为

$$
\begin{aligned}
\min \quad & x_7 + x_8 + x_9 ; \\
\text{s.t.} \quad & 2x_6 + x_7 = 1, \\
& x_6 + x_8 = 3, \\
& x_6 + x_9 = 2, \\
& x_6, x_7, x_8, x_9 \geqslant 0.
\end{aligned}
$$

用单纯形法求得其最优解为

$$
(x_6, x_7, x_8, x_9) = \left(\frac{1}{2}, 0, \frac{5}{2}, \frac{3}{2}\right),
$$

最优目标值为 4. 其对偶问题为

$$\begin{aligned}
\max \quad & v_1 + 3v_2 + 2v_3; \\
\text{s.t.} \quad & 2v_1 + v_2 + v_3 \leqslant 0, \\
& v_1 \leqslant 1, \\
& v_2 \leqslant 1, \\
& v_3 \leqslant 1.
\end{aligned}$$

由于限定原问题对解$\left(\frac{1}{2},0,\frac{5}{2},\frac{3}{2}\right)$来说，约束 $x_8 \geqslant 0$、$x_9 \geqslant 0$ 不等式成立，故对偶问题中 $v_2 \leqslant 1$、$v_3 \leqslant 1$ 应该等式成立，于是得最优解 $\boldsymbol{v}=(-1,1,1)$. 对于 $j=1,2,3,4,5$ 计算 $\boldsymbol{va}_j$ 得 $\boldsymbol{va}_1=0$，$\boldsymbol{va}_2=4$，$\boldsymbol{va}_3=-2$，$\boldsymbol{va}_4=0$，$\boldsymbol{va}_5=-4$. 于是按照(1.62)求得，$\theta=1/2$，从而得新的对偶可行解为

$$\boldsymbol{w} = (0,0,0) + \frac{1}{2}(-1,1,1) = \left(-\frac{1}{2},\frac{1}{2},\frac{1}{2}\right),$$

相应地，$Q=\{2,6\}$. 新的限定原问题为

$$\begin{aligned}
\min \quad & x_7 + x_8 + x_9; \\
\text{s.t.} \quad & -x_2 + 2x_6 + x_7 = 1, \\
& 2x_2 + x_6 + x_8 = 3, \\
& x_2 + x_6 + x_9 = 2, \\
& x_2, x_6, x_7, x_8, x_9 \geqslant 0.
\end{aligned}$$

最优解为

$$(x_2, x_6, x_7, x_8, x_9) = (1,1,0,0,0),$$

目标函数最优值为 0. 这样，便求得原问题的最优解为

$$x^* = (0,1,0,0,0,1)^{\mathrm{T}},$$

目标函数最优值为 2.

### 1.5.6 原始-对偶单纯形表

原始-对偶单纯形法的全部运算也可以在表上完成.

**表 1.43　原始-对偶单纯形表**

| | $x$ | $x_a$ | |
|---|---|---|---|
| $x_B(=x_a)$ | $A_{m\times n}$ | $I_m$ | $b$ |
| $g$ | $0$ | $-e$ | $0$ |
| $f$ | $wA-c$ | $0$ | $wb$ |

表 1.43 中有两个目标行：第 $m+1$ 行在迭代过程中存放限定原问题的检验数 $\hat{z}_j-\hat{c}_j=va_j-\hat{c}_j$，右边存放相应的目标值；第 $m+2$ 行是对偶可行解 $w$ 所对应的 $wa_j-c_j$，即检验数 $z_j-c_j$.

运算过程中，用空白方框□标明限定原问题中的变量．即使得 $z_j-c_j=0$ 的那些 $x_j$，在求解限定原问题时，只有这些变量可以进基．并且，仅对表中前面 $m+1$ 行进行运算．又因为对原来的变量 $x_j$ 来说，$\hat{c}_j=0$，故求解完一个限定原问题后，要修改对偶可行解时，可按下式

$$\theta=\min\left\{-\frac{wa_j-c_j}{\hat{z}_j-\hat{c}_j}\middle|\hat{z}_j-\hat{c}_j>0\right\}\qquad(1.63)$$

求出 $\theta$．然后，将第 $m+1$ 行乘以 $\theta$ 后加到第 $m+2$ 行上，即可以更换限定原问题．

计算步骤：

(1) 找出对偶初始可行解 $w$，建立原始-对偶单纯形表 1.43．通过初等行变换，将第 $m+1$ 行中 $x_a$ 对应的检验数变为 0.

(2) 令 $Q=\{j\mid wa_j-c_j=0\}$．对于 $j\in Q$，在相应的 $x_j$ 上面标明□.

(3) 求 $\hat{z}_k-\hat{c}_k=\max\{\hat{z}_j-\hat{c}_j\mid j\in Q\}$.

(4) 若 $\hat{z}_k-\hat{c}_k>0$，则 $x_k$ 为进基变量．按最小比原则确定离基变量，进行消元法运算，返回(3)；若 $\hat{z}_k-\hat{c}_k\leqslant 0$，且 $g_0=0$，停止计算，已求得原问题的最优解；否则，转(5).

(5) 若 $\hat{z}_j-\hat{c}_j\leqslant 0, j\in Q$，且 $g_0>0$，停止计算，原问题无解；否则，按(1.63)求 $\theta$，将第 $m+1$ 行乘以 $\theta$ 后加到第 $m+2$ 行上去，返回(2).

**例 1.5.15** 用原始-对偶单纯形表求解例 1.5.14

**解** 取 $w=(0,0,0)$ 作为对偶问题的初始解. 建立初始原始-对偶单纯形表 1.44.

**表 1.44**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_7$ | 1 | −1 | 2 | −1 | 1 | 2 | 1 | 0 | 0 | 1 |
| $x_8$ | −1 | 2 | 1 | −2 | −1 | 1 | 0 | 1 | 0 | 3 |
| $x_9$ | 2 | 1 | −1 | 1 | −2 | 1 | 0 | 0 | 1 | 2 |
| $g$ | 0 | 0 | 0 | 0 | 0 | 0 | −1 | −1 | −1 | 0 |
| $f$ | −3 | −2 | −1 | −2 | −2 | 0 | 0 | 0 | 0 | 0 |

将第 1,2,3 行加到第 4 行上去，从而将 $x_7,x_8,x_9$ 的检验数消为 0，得表 1.45.

**表 1.45**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ □ | $x_7$ □ | $x_8$ □ | $x_9$ □ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_7$ | 1 | −1 | 2 | −1 | 1 | [2] | 1 | 0 | 0 | 1 |
| $x_8$ | −1 | 2 | 1 | −2 | −1 | 1 | 0 | 1 | 0 | 3 |
| $x_9$ | 2 | 1 | −1 | 1 | −2 | 1 | 0 | 0 | 1 | 2 |
| $g$ | 2 | 2 | 2 | −2 | −2 | 4 | 0 | 0 | 0 | 6 |
| $f$ | −3 | −2 | −1 | −2 | −2 | 0 | 0 | 0 | 0 | 0 |

由于 $Q=\{6\}$，故当前限定原问题的变量为 $x_6$ 和人工变量 $x_7$，$x_8$，$x_9$. 又因为 $\hat{z}_6-\hat{c}_6=4=\max\{\hat{z}_j-\hat{c}_j \mid j\in Q\}$，故 $x_6$ 为进基变量.

按最小比确定 $x_7$ 为离基变量. 通过消元法运算得表 1.46.

**表 1.46**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_6$ | $\frac{1}{2}$ | $-\frac{1}{2}$ | 1 | $-\frac{1}{2}$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ | 0 | 0 | $\frac{1}{2}$ |
| $x_8$ | $-\frac{3}{2}$ | $\frac{5}{2}$ | 0 | $-\frac{3}{2}$ | $-\frac{3}{2}$ | 0 | $-\frac{1}{2}$ | 1 | 0 | $\frac{5}{2}$ |
| $x_9$ | $\frac{3}{2}$ | $\frac{3}{2}$ | $-2$ | $\frac{3}{2}$ | $-\frac{5}{2}$ | 0 | $-\frac{1}{2}$ | 0 | 1 | $\frac{3}{2}$ |
| $g$ | 0 | 4 | $-2$ | 0 | $-4$ | 0 | $-2$ | 0 | 0 | 4 |
| $f$ | $-3$ | $-2$ | $-1$ | $-2$ | $-2$ | 0 | 0 | 0 | 0 | 0 |

由于对限定原问题中所有的变量来说, $\hat{z}_j-\hat{c}_j\leqslant 0$, 故已找到了限定原问题的最优解. 按(1.63)求得 $\theta=1/2$. 将第 4 行乘以 1/2 加到第 5 行上去, 得表 1.47.

**表 1.47**

| | $x_1$ | $x_2$ □ | $x_3$ | $x_4$ | $x_5$ | $x_6$ □ | $x_7$ | $x_8$ □ | $x_9$ □ | |
|---|---|---|---|---|---|---|---|---|---|---|
| $x_6$ | $\frac{1}{2}$ | $-\frac{1}{2}$ | 1 | $-\frac{1}{2}$ | $\frac{1}{2}$ | 1 | $\frac{1}{2}$ | 0 | 0 | $\frac{1}{2}$ |
| $x_8$ | $-\frac{3}{2}$ | $\frac{5}{2}$ | 0 | $-\frac{3}{2}$ | $-\frac{3}{2}$ | 0 | $-\frac{1}{2}$ | 1 | 0 | $\frac{5}{2}$ |
| $x_9$ | $\frac{3}{2}$ | $\frac{3}{2}$ | $-2$ | $\frac{3}{2}$ | $-\frac{5}{2}$ | 0 | $-\frac{1}{2}$ | 0 | 1 | $\frac{3}{2}$ |
| $g$ | 0 | 4 | $-2$ | 0 | $-4$ | 0 | $-2$ | 0 | 0 | 4 |
| $f$ | $-3$ | 0 | $-2$ | $-2$ | $-4$ | 0 | $-1$ | 0 | 0 | 2 |

由于 $\hat{z}_2-\hat{c}_2=4>0$, 故 $x_2$ 为进基变量. 并取 $x_8$ 为离基变量, 进行消元法运算得表 1.48.

表 1.48

| | | □ | | | | □ | | □ | □ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | |
| $x_6$ | $\frac{1}{5}$ | 0 | 1 | $-\frac{4}{5}$ | $\frac{1}{5}$ | 1 | $\frac{2}{5}$ | $\frac{1}{5}$ | 0 | 1 |
| $x_2$ | $-\frac{3}{5}$ | 1 | 0 | $-\frac{3}{5}$ | $-\frac{3}{5}$ | 0 | $-\frac{1}{5}$ | $\frac{2}{5}$ | 0 | 1 |
| $x_9$ | $\frac{12}{5}$ | 0 | $-2$ | $\frac{12}{5}$ | $-\frac{8}{5}$ | 0 | $-\frac{1}{5}$ | $-\frac{3}{5}$ | 1 | 0 |
| $g$ | $\frac{12}{5}$ | 0 | $-2$ | $\frac{12}{5}$ | $-\frac{8}{5}$ | 0 | $-\frac{6}{5}$ | $-\frac{8}{5}$ | 0 | 0 |
| $f$ | $-3$ | 0 | $-2$ | $-2$ | $-4$ | 0 | $-1$ | 0 | 0 | 2 |

在表 1.48 中由于限定原问题的所有检验数 $\hat{z}_j-\hat{c}_j\leqslant 0$，并且 $g_0=0$，故已经找到了原问题的最优解 $\boldsymbol{x}^*=(0,1,0,0,0,1)^{\mathrm{T}}$，目标函数最优值为 2.

### 1.5.7 寻找对偶初始解的人工约束法

对偶单纯形法要求有一个对偶基本可行解作为初始解，原始-对偶单纯形法要求有一个对偶可行解作为初始解. 对于极小化问题，如果原问题的费用系数向量 $\boldsymbol{c}\leqslant\boldsymbol{0}$，则 $\boldsymbol{w}\leqslant\boldsymbol{0}$ 显然是一个对偶可行解. 当对偶问题没有明显的基本可行解或可行解时，可采用**人工约束法**(artificial constraint technique).

设线性规划问题

$$
\begin{aligned}
(\mathrm{P})\quad \min\ & \boldsymbol{cx}\ ;\\
\text{s.t.}\ & \boldsymbol{Ax}=\boldsymbol{b},\\
& \boldsymbol{x}\geqslant 0.
\end{aligned}
$$

假设它有初始基 $\boldsymbol{B}=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m)$(因为总可以用高斯消元法找到它的一个基). 将约束方程化为

$$
\boldsymbol{x}_B+\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_N=\boldsymbol{B}^{-1}\boldsymbol{b},
$$

增加一个约束

$$\boldsymbol{e}\boldsymbol{x}_N \leqslant M,$$

这里,$M$ 是很大的正数,$\boldsymbol{e}$ 是 $n-m$ 维行向量.由于 $M$ 很大,这个新增加的约束对原问题不发生影响.增加新约束后的问题称为**扩充问题**.引入松弛变量 $x_{n+1}$,得到扩充问题的单纯形表为

表 1.49

| | $\boldsymbol{x}_B$ | $\boldsymbol{x}_N$ | $x_{n+1}$ | |
|---|---|---|---|---|
| $\boldsymbol{x}_B$ | $\boldsymbol{I}$ | $\boldsymbol{N}$ | 0 | $\bar{\boldsymbol{b}}$ |
| $x_{n+1}$ | $\boldsymbol{0}$ | $\boldsymbol{e}$ | 1 | $M$ |
| | $\boldsymbol{0}$ | $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{N}-\boldsymbol{c}_N$ | 0 | $f$ |

设 $z_k-c_k=\max\{z_j-c_j\}$.选 $x_k$ 为进基变量,$x_{n+1}$ 离基.以 $\bar{a}_{m+1,k}$ 为主元,进行消元法运算得新表.特别,将附加行乘以$[-(z_k-c_k)]$加到目标行上去.由 $k$ 的选取方法和消元法运算,新表中必有 $z_j-c_j\leqslant 0$ 对一切 $j$ 成立.因此,找到一个对偶基本可行解.于是,便可以由此开始用对偶单纯形法或原始-对偶单纯形法求解扩充问题.

计算结果有两种情况:

(1) 扩充问题无解.这时,原问题也无解.

(2) 扩充问题有最优解

$$\bar{\boldsymbol{x}}=(\bar{x}_1,\cdots,\bar{x}_n)=\bar{\boldsymbol{x}}^{(1)}+M\bar{\boldsymbol{x}}^{(2)}$$

最优值为

$f_0=f_1+Mf_2$.由于 $M$ 很大,故必有 $f_2\leqslant 0$.

1) 若 $f_2<0$,这时必有 $\bar{\boldsymbol{x}}^{(2)}\neq 0$.当 $M\to+\infty$时,$\bar{\boldsymbol{x}}$ 总是原问题的可行解.此时,$f_0\to-\infty$,故原问题有无界解.

2) 若 $f_2=0$,这时 $f_0=f_1$ 是原问题的最优值,$\bar{\boldsymbol{x}}\geqslant\boldsymbol{0}$ 为原问题的最优解.若 $\bar{\boldsymbol{x}}^{(2)}=\boldsymbol{0}$,$\bar{\boldsymbol{x}}$ 还是一个基本可行解.否则,令 $M_0=$

$\min\{M|\bar{x}^{(1)}+M\bar{x}^{(2)}\geqslant 0\}$，则

$$\bar{x}^0=\bar{x}^{(1)}+M_0\bar{x}^{(2)}$$

也是基本可行解.

**例 1.5.16** 用对偶单纯形法求解线性规划

$$\begin{aligned}\min\quad & -2x_4+4x_5\;;\\ \text{s.t.}\quad & x_1+x_4-2x_5=2,\\ & x_2-3x_4+x_5=3,\\ & x_3-x_4-x_5=-2,\\ & x_1,x_2,\cdots,x_5\geqslant 0.\end{aligned}$$

**解** 该问题有一个明显的基本解$(2,3,-2,0,0)^{\mathrm{T}}$，但它不是正则解. 增加一个约束 $x_4+x_5\leqslant M$，并引入松弛变量 $x_6$，得扩充问题的单纯形表 1.50.

**表 1.50**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 1 | $-2$ | 0 | 2 |
| $x_2$ | 0 | 1 | 0 | $-3$ | 1 | 0 | 3 |
| $x_3$ | 0 | 0 | 1 | $-1$ | $-1$ | 0 | $-2$ |
| $x_6$ | 0 | 0 | 0 | [1] | 1 | 1 | $M$ |
| $f$ | 0 | 0 | 0 | 2 | $-4$ | 0 | 0 |

因为 $z_4-c_4=2>0$，故 $x_4$ 进基，$x_6$ 离基，以 $\bar{a}_{44}$ 为主元，进行消元法运算，得表 1.51.

**表 1.51**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 0 | [$-3$] | $-1$ | $2-M$ |
| $x_2$ | 0 | 1 | 0 | 0 | 4 | 3 | $3+3M$ |

续表

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | $-2+M$ |
| $x_4$ | 0 | 0 | 0 | 1 | 1 | 1 | $M$ |
| $f$ | 0 | 0 | 0 | 0 | $-6$ | $-2$ | $-2M$ |

表 1.51 对应着扩充问题的一个正则解. 下面便可以由这张表开始用对偶单纯形法求解扩充问题. 以 $\bar{a}_{15}=-3$ 为主元,进行消元法运算,得表 1.52.

**表 1.52**

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_5$ | $-\frac{1}{3}$ | 0 | 0 | 0 | 1 | $\frac{1}{3}$ | $-\frac{2}{3}+\frac{1}{3}M$ |
| $x_2$ | $\frac{4}{3}$ | 1 | 0 | 0 | 0 | $\frac{5}{3}$ | $\frac{17}{3}+\frac{5}{3}M$ |
| $x_3$ | 0 | 0 | 1 | 0 | 0 | 1 | $-2+M$ |
| $x_4$ | $\frac{1}{3}$ | 0 | 0 | 1 | 0 | $\frac{2}{3}$ | $\frac{2}{3}+\frac{2}{3}M$ |
| $f$ | $-2$ | 0 | 0 | 0 | 0 | 0 | $-4$ |

由于这张表的 $\bar{\boldsymbol{b}}\geqslant\boldsymbol{0}$,因此得到扩充问题的最优解

$$\bar{\boldsymbol{x}}=\left(0,\frac{17}{3}+\frac{5}{3}M,-2+M,\frac{2}{3}+\frac{2}{3}M,-\frac{2}{3}+\frac{1}{3}M,0\right)^{\mathrm{T}}$$

目标最优值为 $f_0=-4$.

故原问题有最优解. 容易求得 $M_0=2$,得原问题的最优基本可行解为

$$\boldsymbol{x}=(0,9,0,2,0)^{\mathrm{T}},$$

目标函数最优值仍为 $f_0=-4$.

## 1.6 分解算法

实际中遇到的线性规划问题由于约束矩阵太大，即使应用计算机求解也很困难.但是，有不少问题的约束具有特定的结构，因而可以设法将它们分解为若干个较小的线性规划问题再来求解.这类方法称为**分解算法**(decomposition algorithm).现介绍Dantzig-Wolfe 分解方法，简称 D-W 分解方法.

### 1.6.1 D-W 分解方法概述

设下列线性规划

$$\begin{aligned} \min \quad & \boldsymbol{cx}; \\ \text{s.t.} \quad & \boldsymbol{Ax} = \boldsymbol{b}, \\ & \boldsymbol{x} \in X. \end{aligned} \tag{1.64}$$

其中，$\boldsymbol{A}$ 为 $m\times m$ 矩阵，$\boldsymbol{b}$ 为 $m$ 维列向量，$X$ 表示有特定结构的多面集，并且假设 $X$ 是有界多面集.于是，$X$ 只有有限个极点：$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\cdots$，$\boldsymbol{x}_t$.这样，任何 $\boldsymbol{x}\in X$，可以表示为这些极点的凸组合，即

$$\begin{cases} \boldsymbol{x} = \sum\limits_{j=1}^{t} \lambda_j \boldsymbol{x}_j, \\ \sum\limits_{j=1}^{t} \lambda_j = 1,\ \lambda_j \geqslant 0,\ j = 1,2,\cdots,t. \end{cases} \tag{1.65}$$

如果用(1.65)代替 $\boldsymbol{x}$，则(1.64)改写成以 $\lambda_1$，$\lambda_2$，$\cdots$，$\lambda_t$ 为变量的问题：

$$\min \quad \sum_{j=1}^{t} (\boldsymbol{cx}_j)\lambda_j; \tag{1.66}$$

$$\text{s.t.} \quad \sum_{j=1}^{t} (\boldsymbol{Ax}_j)\lambda_j = \boldsymbol{b}, \tag{1.67}$$

$$\sum_{j=1}^{t} \lambda_j = 1,$$

$$\lambda_j \geqslant 0, j = 1,2,\cdots,t. \tag{1.68}$$

称为**主规划**(master program).

由于极点数目 $t$ 很大，因此，在知道 $X$ 的所有极点后直接求解主规划是十分困难的. D-W 方法不要求知道所有的极点就能求解主规划，从而求得原问题的最优解. 它是应用修正单纯形法来实现的.

设考虑用修正单纯形法解主规划(1.66)~(1.68). 设初始基为 $\boldsymbol{B}$，它是 $m+1$ 阶方阵. 分别用 $\boldsymbol{w}$ 和 $\boldsymbol{v}$ 表示相应于(1.67)和(1.68)的对偶变量；则 $(\boldsymbol{w},\boldsymbol{v})=\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}$，其中 $\hat{\boldsymbol{c}}_B$ 是由基变量的费用系数组成的行向量. 再记 $\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\begin{bmatrix}\boldsymbol{b}\\1\end{bmatrix}$. 将修正单纯形法数据列成下表 1.53.

表 1.53

| 基 | 右端 |
|---|---|
| $(\boldsymbol{w},\boldsymbol{v})$ | $\hat{\boldsymbol{c}}_B\bar{\boldsymbol{b}}$ |
| $\boldsymbol{B}^{-1}$ | $\bar{\boldsymbol{b}}$ |

现在要检验当前解的最优性. 为此，首先求出最大检验数

$$\begin{aligned} z_k - \hat{c}_k &= \max_{1\leqslant j\leqslant t}\{z_j - \hat{c}_j\} \\ &= \max_{1\leqslant j\leqslant t}\left\{(\boldsymbol{w},\boldsymbol{v})\begin{bmatrix}\boldsymbol{A}\boldsymbol{x}_j\\1\end{bmatrix} - \boldsymbol{c}\boldsymbol{x}_j\right\} \\ &= \max_{1\leqslant j\leqslant t}\{\boldsymbol{w}\boldsymbol{A}\boldsymbol{x}_j + \boldsymbol{v} - \boldsymbol{c}\boldsymbol{x}_j\}. \end{aligned} \tag{1.69}$$

由于不能事先求出所有的极点 $\boldsymbol{x}_j$，又由于线性目标函数在有界多面集上的最大值必在它的某个极点处达到，因此，(1.69)的值恰好等于下列线性规划问题的最优目标值.

$$\begin{aligned} \max \quad & (\boldsymbol{wA} - \boldsymbol{c})\boldsymbol{x} + v; \\ \text{s.t.} \quad & \boldsymbol{x} \in X. \end{aligned} \tag{1.70}$$

(1.70)称为**子问题**(subproblem). 由于 $X$ 具有特定的结构,子问题是较容易求解的. 设用单纯形法求得子问题的最优解为 $\boldsymbol{x}_k$,最优值为 $z_k - \hat{c}_k$.

如果 $z_k - \hat{c}_k = 0$,则主规划的当前基本可行解是最优解,从而可求出(1.64)的最优解.

如果 $z_k - \hat{c}_k > 0$,则相应的非基变量 $\lambda_k$ 进基. 计算主列 $\bar{\boldsymbol{a}}_k = \boldsymbol{B}^{-1}\begin{bmatrix} \boldsymbol{Ax}_k \\ 1 \end{bmatrix}$,并且将 $\begin{bmatrix} z_k - \hat{c}_k \\ \bar{\boldsymbol{a}}_k \end{bmatrix}$ 附在修正单纯形表右端. 按最小比值规则确定离基变量 $\lambda_{B_r}$. 以 $\bar{a}_{rk}$ 为主元,进行消元法运算得新表. 这样完成了 D-W 方法的一次迭代. 如此反复进行,直至求得主规划的最优解.

### 1.6.2 D-W 分解方法的计算步骤

(1) 写出主规划. 找出主规划的一个初始基本可行解,对应的初始基为 $\boldsymbol{B}$. 计算

$$(\boldsymbol{w}\ v) = \hat{\boldsymbol{c}}_B \boldsymbol{B}^{-1},$$

$$\bar{\boldsymbol{b}} = \boldsymbol{B}^{-1}\begin{bmatrix} \boldsymbol{b} \\ 1 \end{bmatrix}.$$

构造初始修正单纯形表 1.53.

(2) 解子问题(1.70). 设最优解为 $\boldsymbol{x}_k$,最优值为 $z_k - \hat{c}_k$. 如果 $z_k - \hat{c}_k = 0$,停止计算. 主规划的当前基本可行解为最优解. 计算(1.64)的最优解

$$\boldsymbol{x} = \sum_{j \in J_B} \lambda_j \boldsymbol{x}_j,$$

其中,$J_B$ 是主规划的最优解中基变量的指标集. 如果 $z_k - \hat{c}_k > 0$,转(3).

(3) 计算主列. 令

$$\bar{\boldsymbol{a}}_k = \boldsymbol{B}^{-1}\begin{bmatrix} \boldsymbol{A}\boldsymbol{x}_k \\ 1 \end{bmatrix},$$

组成下表 1.54

**表 1.54**

| | | | | $\lambda_k$ |
|---|---|---|---|---|
| | $(\boldsymbol{w},\boldsymbol{v})$ | $\boldsymbol{c}_B\bar{\boldsymbol{b}}$ | | $z_k - c_k$ |
| $\Lambda_B$ | $\boldsymbol{B}^{-1}$ | $\bar{\boldsymbol{b}}$ | | $\bar{\boldsymbol{a}}_k$ |

计算

$$\frac{\bar{b}_r}{\bar{a}_{rk}} = \min\left\{\frac{\bar{b}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik} > 0\right\}.$$

以 $\bar{a}_{rk}$ 为主元，进行消元法运算. 删去表 1.54 最右边一列，返回 (2).

**例 1.6.1** 用分解算法解下列问题

$$\begin{aligned}
\min \quad & -x_1 - 3x_2 + x_3 - x_4; \\
\text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 \leqslant 8, \\
& x_1 + x_2 \leqslant 6, \\
& x_3 + 2x_4 \leqslant 10, \\
& -x_3 + x_4 \leqslant 4, \\
& x_1, x_2, x_3, x_4 \geqslant 0.
\end{aligned}$$

**解** 把约束分为两组：

$$\boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b} \text{ 和 } \boldsymbol{x} \in X,$$

其中

$$\begin{aligned}
\boldsymbol{A} &= (1,1,1,1), \\
\boldsymbol{b} &= 8, \\
X &= X_1 \times X_2,
\end{aligned}$$

$$X_1=\{(x_1,x_2)\mid x_1+x_2\leqslant 6,(x_1,x_2)\geqslant \mathbf{0}\},$$

$$X_2=\left\{(x_3,x_4)\left|\begin{array}{l}x_3+2x_4\leqslant 10\\ -x_3+\ \ x_4\leqslant 4\\ (x_3,x_4)\geqslant \mathbf{0}\end{array}\right.\right\},$$

并记 $\boldsymbol{c}=(-1,-3,1,-1)$ 为费用系数向量. 引进松弛变量 $s$ 得主规划:

$$\begin{aligned}
\min \quad & \sum_{j=1}^{t}\hat{c}_j\lambda_j;\\
\text{s.t.} \quad & s+\sum_{j=1}^{t}(\boldsymbol{A}\boldsymbol{x}_j)\lambda_j=8,\\
& \sum_{j=1}^{t}\lambda_j=1,\\
& \lambda_j\geqslant 0,\ j=1,\cdots,t,\\
& s\geqslant 0.
\end{aligned}$$

其中,$\boldsymbol{x}_j$ 为 $X$ 的极点,$\hat{c}_j=\boldsymbol{c}\boldsymbol{x}_j$.

为了找出主规划的一个初始基本可行解,作出 $X_1$ 和 $X_2$ 的图 1.3.

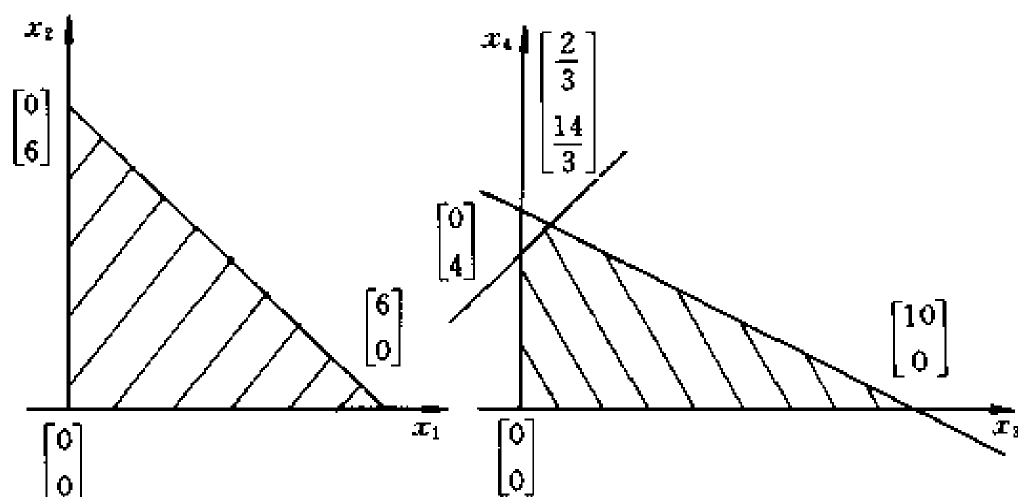



图 1.3

求得 $\boldsymbol{x}_1=(0,0,0,0)^{\mathrm{T}}$ 是 $X$ 的一个极点，故可取 $s$ 和 $\lambda_1$ 为初始基变量，于是

$$\boldsymbol{B}=\begin{bmatrix}1 & \boldsymbol{A}\boldsymbol{x}_1\\ 0 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}=\boldsymbol{B}^{-1},$$

$$\hat{\boldsymbol{c}}_B=(0,\boldsymbol{c}\boldsymbol{x}_1)=(0,0),$$

$$(\boldsymbol{w},\boldsymbol{v})=\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}=(0,0),$$

$$\bar{\boldsymbol{b}}=\boldsymbol{B}^{-1}\begin{bmatrix}\boldsymbol{b}\\ 1\end{bmatrix}=\begin{bmatrix}8\\ 1\end{bmatrix},$$

$$\hat{\boldsymbol{c}}_B\bar{\boldsymbol{b}}=0.$$

构造主规划的初始修正单纯形表为

| | | | |
|---|---|---|---|
| | 0 | 0 | 0 |
| $s$ | 1 | 0 | 8 |
| $\lambda_1$ | 0 | 1 | 1 |

**迭代 1**　子问题为

$$\begin{aligned}&\max\quad (\boldsymbol{w}\boldsymbol{A}-\boldsymbol{c})\boldsymbol{x}+\boldsymbol{v}\\ =&\max\ -\boldsymbol{c}\boldsymbol{x}\\ =&\max\quad x_1+3x_2-x_3+x_4;\\ &\text{s.t.}\quad \boldsymbol{x}\in X.\end{aligned}$$

利用单纯形法求得子问题的最优解

$$\boldsymbol{x}_2=(0,6,0,4)^{\mathrm{T}},$$

目标函数最优值为 22. 因此，主规划的最大检验数为

$$z_2-\hat{c}_2=22.$$

计算主列

$$\bar{\boldsymbol{a}}_2=\boldsymbol{B}^{-1}\begin{bmatrix}\boldsymbol{A}\boldsymbol{x}_2\\ 1\end{bmatrix}=\begin{bmatrix}10\\ 1\end{bmatrix}.$$

组成下表

|  | 0 | 0 | 0 | 22 |
|---|---|---|---|---|
| $s$ | 1 | 0 | 8 | $\boxed{10}$ |
| $\lambda_1$ | 0 | 1 | 1 | 1 |

计算 $\min\left\{\dfrac{\bar{b}_1}{\bar{a}_{12}},\dfrac{\bar{b}_2}{\bar{a}_{22}}\right\}=\min\left\{\dfrac{8}{10},\dfrac{1}{1}\right\}=\dfrac{8}{10}$,故 $r=1$,松弛变量 $s$ 离基.以 $\bar{a}_{12}=10$ 为主元,进行消元法运算得新表

|  | $-\dfrac{11}{5}$ | 0 | $-\dfrac{44}{5}$ |
|---|---|---|---|
| $\lambda_2$ | $\dfrac{1}{10}$ | 0 | $\dfrac{2}{5}$ |
| $\lambda_1$ | $-\dfrac{1}{10}$ | 1 | $\dfrac{3}{5}$ |

当前迭代点为

$$\boldsymbol{x}=\lambda_1\boldsymbol{x}_1+\lambda_2\boldsymbol{x}_2=\frac{2}{5}(0,6,0,4)^{\mathrm{T}},$$

目标函数值为$-44/5$,与原来的值 0 相比是减小了.

**迭代 2**

由现在的$(\boldsymbol{w},\boldsymbol{v})=\left(-\dfrac{11}{5},0\right)$得子问题为

$$\begin{aligned}&\max \quad -\frac{6}{5}x_1+\frac{4}{5}x_2-\frac{16}{5}x_3-\frac{6}{5}x_4;\\&\text{s.t.} \quad \boldsymbol{x}\in X.\end{aligned}$$

求得其最优解

$$\boldsymbol{x}_3=(0,6,0,0)^{\mathrm{T}},$$

最优值为$\dfrac{24}{5}$.这时,松弛变量 $s$ 的检验数为

$$(\boldsymbol{w},\boldsymbol{v})\begin{bmatrix}1\\0\end{bmatrix}-0=\boldsymbol{w}=-\frac{11}{5}.$$

因此,最大检验数为 $z_3-\hat{c}_3=\frac{24}{5}$. 计算主列

$$\bar{\boldsymbol{a}}_3=\boldsymbol{B}^{-1}\begin{bmatrix}\boldsymbol{A}\boldsymbol{x}_3\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{10} & 0\\ -\frac{1}{10} & 1\end{bmatrix}\begin{bmatrix}6\\1\end{bmatrix}=\begin{bmatrix}\frac{3}{5}\\ \frac{2}{5}\end{bmatrix}.$$

构造下表

| | | | | |
|---|---|---|---|---|
| | $-\frac{11}{5}$ | 0 | $-\frac{44}{5}$ | $\frac{24}{5}$ |
| $\lambda_2$ | $\frac{1}{10}$ | 0 | $\frac{2}{5}$ | $\boxed{\frac{3}{5}}$ |
| $\lambda_1$ | $-\frac{1}{10}$ | 1 | $\frac{3}{5}$ | $\frac{2}{5}$ |

计算 $\min\left\{\frac{\bar{b}_1}{\bar{a}_{13}},\frac{\bar{b}_2}{\bar{a}_{23}}\right\}=\min\left\{\frac{2}{5}\Big/\frac{3}{5},\frac{3}{5}\Big/\frac{2}{5}\right\}=\frac{2}{5}\Big/\frac{3}{5}$,故 $r=1$,$\lambda_2$ 为离基变量. 以 $\bar{a}_{13}$为主元,进行消元法运算得下表

| | | |
|---|---|---|
| $-3$ | 0 | $-12$ |
| $\frac{1}{6}$ | 0 | $\frac{2}{3}$ |
| $-\frac{1}{6}$ | 1 | $\frac{7}{15}$ |

于是得主规划的基本可行解为

$$\boldsymbol{x}=\lambda_1\boldsymbol{x}_1+\lambda_3\boldsymbol{x}_3=(0,4,0,0)^{\mathrm{T}},$$

目标函数值为$-12$.

**迭代 3**

由当前的$(\boldsymbol{w},\boldsymbol{v})=(-3,0)$得子问题为

$$\max\quad -2x_1\quad -4x_3-2x_4;$$

$$\text{s.t.} \quad x \in X.$$

用单纯形法求得其最优解

$$x_4 = (0,0,0,0)^T,$$

最优目标值为 0. 这时,松弛变量 $s$ 的检验数为 $-3<0$.

本结果表明,最优性检验已通过. 即第二次迭代所得结果已经是最优解了. 因此,原问题的最优解是

$$x = (0,4,0,0)^T,$$

目标函数最优值为 $-12$.

## 1.7 灵敏度分析

在实际应用中,许多线性规划问题的数据不能精确地知道,通常是根据经验估计或用预测方法得到. 因此,某些数据可能需要修改;有时,还可能增加新变量或新约束. 遇到上述情况时,要考虑某些条件变化对最优解的影响,并不需要从头开始计算,只须修改原来问题的最优单纯形表中相应的部分,以便得到条件变化后新问题的单纯形表,再判断所得到的解是否还是新问题的最优解. 如果不是,可继续迭代求解. **灵敏度分析**(sensitivity analysis)就是研究这些问题的.

设如下的线性规划问题

$$\begin{aligned} \min \quad & cx \;; \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \tag{1.71}$$

假定已经求得它的最优单纯形表为表 1.55.

表 1.55

| | $x_B$ | $x_N$ | 右端 |
|---|---|---|---|
| $x_B$ | $I$ | $B^{-1}N$ | $B^{-1}b$ |
| $f$ | $0$ | $C_B B^{-1}N - C_N$ | $C_B B^{-1}b$ |

下面叙述各种变化时的灵敏度分析.

### 1.7.1 改变费用系数向量

假定费用系数向量 $\boldsymbol{c}$ 变成 $\hat{\boldsymbol{c}}$、这时,可行域一般不变化. 故原问题的最优解还是新问题的基本可行解. 但是,需要修改目标行. 新检验数为 $\hat{z}_j-\hat{c}_j=\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}\boldsymbol{a}_j-\hat{c}_j$,新目标函数值为 $\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}\boldsymbol{b}$. 然后,根据 $\hat{z}_j-\hat{c}_j\leqslant 0$ 是否满足,决定是停止计算还是继续迭代.

特别,当费用系数向量只有一个分量 $c_k$ 变成 $\hat{c}_k$ 时,可按下法处理.

(1) 对应的 $x_k$ 是非基变量

这时只需修改检验数 $z_k-c_k$. 由于 $\boldsymbol{c}_B$ 不变化,故新检验数为

$$z_k-\hat{c}_k=z_k-c_k+(c_k-\hat{c}_k),$$

这样就得到了新问题的单纯形表. 如果 $z_k-\hat{c}_k\leqslant 0$,则所得到的解还是新问题的最优解;否则,由此开始继续迭代.

(2) 对应的 $x_k$ 是基变量

设 $x_k=x_{B_t}$. 这时,$c_{B_t}$变成 $\hat{c}_{B_t}$ 新检验数中仍有基变量对应的检验数为 0,而非基变量对应的检验数为

$$\begin{aligned}\hat{z}_j-c_j&=\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}\boldsymbol{a}_j-c_j\\&=\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{a}_j+(0,\cdots,0,\hat{c}_{B_t}-c_{B_t},0,\cdots,0)\boldsymbol{B}^{-1}\boldsymbol{a}_j-c_j\\&=z_j-c_j+(\hat{c}_k-c_k)\bar{a}_{tj}.\end{aligned}\tag{1.72}$$

新目标函数值为

$$\hat{\boldsymbol{c}}_B\boldsymbol{B}^{-1}\boldsymbol{b}=\hat{\boldsymbol{c}}_B\bar{\boldsymbol{b}}=\boldsymbol{c}_B\bar{\boldsymbol{b}}+(\hat{c}_k-c_k)\bar{b}_t.\tag{1.73}$$

故只需将第 $t$ 行乘以$(\hat{c}_k-c_k)$加到目标行上去,再令 $x_k$ 对应的检验数 $\hat{z}_k-\hat{c}_k=0$,便可得到新问题的单纯形表.

**例 1.7.1** 给定线性规划问题

$$\begin{aligned} \min \quad & 5x_1 - 4x_2 \ ; \\ \text{s.t.} \quad & x_1 + x_2 \leqslant 4, \\ & -x_1 + x_2 \leqslant 2, \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

已经求得它的最优单纯形表为

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{b}$ |
|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | −1 | 2 |
| $x_2$ | −1 | 1 | 0 | 1 | 2 |
| $f$ | −1 | 0 | 0 | −4 | −8 |

试考虑下列问题的处理方法.

(1) 若 $c_1=5$ 变成 $\hat{c}_1=-1$,求新问题的最优解.

(2) 若 $c_2=-4$ 变成 $\hat{c}_2=-2$,求新问题的最优解.

**解** (1) 由于 $x_1$ 是非基变量,故只需计算 $z_1-\hat{c}_1$:

$$z_1-\hat{c}_1=z_1-c_1+(c_1-\hat{c}_1)=-1+6=5.$$

将所给最优表中 $x_1$ 对应的检验数改成 5,得新问题的单纯形表如下:

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{b}$ |
|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | −1 | 2 |
| $x_2$ | −1 | 1 | 0 | 1 | 2 |
| $f$ | 5 | 0 | 0 | −4 | −8 |

由于 $z_1-\hat{c}_1>0$,故需由此表开始继续迭代才能求得新问题的最优解.

(2) 当 $c_2=-4$ 变成 $\hat{c}_2=-2$ 时,仍然有 $\hat{z}_2-\hat{c}_2=\hat{z}_3-\hat{c}_3=0$. 按式(1.72)和(1.73)计算 $\hat{z}_1-c_1$,$\hat{z}_4-c_4$ 和新目标值 $f$,得新问

题的单纯形表如下：

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{b}$ |
|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | $-1$ | 2 |
| $x_2$ | $-1$ | 1 | 0 | 1 | 2 |
| $f$ | $-3$ | 0 | 0 | $-2$ | $-4$ |

由于 $\hat{z}_j-c_j\leqslant 0$ 得到满足，所以原问题的最优解 $\boldsymbol{x}=(0,2)^{\mathrm{T}}$ 还是新问题的最优解，而目标函数最优值变成$-4$.

### 1.7.2 改变右端向量

设右端向量 $\boldsymbol{b}$ 变成 $\hat{\boldsymbol{b}}$. 这时，只需修改右端一列便可得到新问题的单纯形表. 新表右端一列为

$$\bar{\hat{\boldsymbol{b}}}=\boldsymbol{B}^{-1}\hat{\boldsymbol{b}},\qquad \hat{f}=\boldsymbol{c}_B\bar{\hat{\boldsymbol{b}}}\ .$$

检验数均不改变，故仍然有 $z_j-c_j\leqslant 0$. 如果 $\bar{\hat{\boldsymbol{b}}}\geqslant\boldsymbol{0}$，则已找到新问题的最优解；否则，单纯形表对应新问题的一个正则解. 于是，可用对偶单纯形法继续求解新问题.

**例 1.7.2** 设例 1.7.1 中的线性规划，若右端向量 $\boldsymbol{b}=\begin{bmatrix}4\\2\end{bmatrix}$变成 $\hat{\boldsymbol{b}}=\begin{bmatrix}3\\5\end{bmatrix}$，求新问题的最优解.

**解** 由于 $\boldsymbol{B}^{-1}=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}$，故

$$\bar{\hat{\boldsymbol{b}}}=\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}\begin{bmatrix}3\\5\end{bmatrix}=\begin{bmatrix}-2\\5\end{bmatrix},$$

$$\hat{f}=(c_3,c_2)\bar{\hat{\boldsymbol{b}}}=(0,-4)\begin{bmatrix}-2\\5\end{bmatrix}=-20.$$

于是得下表

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{\hat{b}}$ |
|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | $-1$ | $-2$ |
| $x_2$ | $-1$ | 1 | 0 | 1 | 5 |
| $\hat{f}$ | $-1$ | 0 | 0 | $-4$ | $-20$ |

此表对应新问题的一个正则解，故由此开始用对偶单纯形法求解新问题.

### 1.7.3 改变矩阵

设矩阵 $\boldsymbol{A}$ 的一列 $\boldsymbol{a}_k$ 变成 $\hat{\boldsymbol{a}}_k$. 分两种情况：

(1) 相应的 $x_k$ 是非基变量

这时，基 $\boldsymbol{B}$ 不变. 只需修改单纯形表中 $x_k$ 对应的一列便可得新问题的单纯形表

$$\bar{\hat{\boldsymbol{a}}}_k = \boldsymbol{B}^{-1}\hat{\boldsymbol{a}}_k,$$

$$\hat{z}_k - c_k = \boldsymbol{c}_B\boldsymbol{B}^{-1}\hat{\boldsymbol{a}}_k - c_k.$$

如果 $\hat{z}_k - c_k \leqslant 0$，则得到的表也是新问题的最优单纯形表；否则，需由此开始继续迭代，求解新问题.

(2) 相应的 $x_k$ 是基变量

设 $x_k = x_{B_l}$. 这时基 $\boldsymbol{B}$ 发生变化，从而单纯形表中的右端列和非基变量对应的每一列都受影响. 为了保持 $\boldsymbol{B}^{-1}$，求解下列辅助问题

$$\begin{aligned}\min \quad & g = x_{n+1};\\ \text{s.t.} \quad & \boldsymbol{a}_1x_1 + \cdots + \hat{\boldsymbol{a}}_kx_k + \cdots + \boldsymbol{a}_nx_n + \boldsymbol{a}_kx_{n+1} = \boldsymbol{b},\\ & x_j \geqslant 0 \quad j = 1,\cdots,n,n+1.\end{aligned} \tag{1.74}$$

如果不考虑 $x_k$ 的系数，则辅助问题的约束矩阵与原问题相同，只是把 $x_k$ 换成了 $x_{n+1}$. 因而，令 $x_{n+1}$ 为基变量，$x_k$ 为非基变量. 即在原问题的最优单纯形表中，将第 $k$ 列移作 $x_{n+1}$对应的列，再令 $\tilde{\boldsymbol{a}}_k=\boldsymbol{B}^{-1}\hat{\boldsymbol{a}}_k$，便可得到辅助问题的一个基本可行解，由此开始解辅助问题. 若最优值 $g=0$，则辅助问题的最优解对应着新问题的一个基本可行解. 从而，可由此开始求解新问题.

**例 1.7.3** 考虑例 1.7.1 中的线性规划. 如果 $\boldsymbol{a}_2=\begin{bmatrix}1\\1\end{bmatrix}$变成 $\hat{\boldsymbol{a}}_2=\begin{bmatrix}-2\\1\end{bmatrix}$，求解新问题.

**解** 由于 $x_2$ 是基变量，故构造如下的辅助问题：

$$
\begin{aligned}
\min \quad & g=x_5;\\
\text{s.t.}\quad & x_1-2x_2+x_3+x_5=4,\\
& -x_1+x_2+x_4+x_5=2,\\
& x_1,\cdots,x_5\geqslant 0.
\end{aligned}
$$

因 $\tilde{\boldsymbol{a}}_2=\boldsymbol{B}^{-1}\hat{\boldsymbol{a}}_2=\begin{bmatrix}1&-1\\0&1\end{bmatrix}\begin{bmatrix}-2\\1\end{bmatrix}=\begin{bmatrix}-3\\1\end{bmatrix}$，故得辅助问题单纯形表如下：

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_3$ | 2 | −3 | 1 | −1 | 0 | 2 |
| $x_5$ | −1 | [1] | 0 | 1 | 1 | 2 |
| $g$ | −1 | 1 | 0 | 1 | 0 | 2 |

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $\bar{b}$ |
|---|---|---|---|---|---|---|
| $x_3$ | −1 | 0 | 1 | 2 | 3 | 8 |
| $x_2$ | −1 | 1 | 0 | 1 | 1 | 2 |
| $g$ | 0 | 0 | 0 | 0 | −1 | 0 |

此时，$g=0$，第一阶段结束. 删去 $g$ 行与 $x_5$ 列，换上 $f$ 行，再进行单纯形法迭代，求解新问题

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $\bar{b}$ |
|---|---|---|---|---|---|
| $x_3$ | $-1$ | 0 | 1 | 2 | 8 |
| $x_2$ | $-1$ | 1 | 0 | 1 | 2 |
| $f$ | $-1$ | 0 | 0 | $-4$ | $-8$ |

由于全部 $z_j-c_j\leqslant 0$，故已求得新问题的最优解为 $\boldsymbol{x}=(0,2)^{\mathrm{T}}$，目标函数最优值为 $\hat{f}=-8$.

### 1.7.4 增加新变量

设增加一个新变量 $x_{n+1}$，其费用系数为 $c_{n+1}$，在约束矩阵中对应的向量为 $\boldsymbol{a}_{n+1}$. 则把 $x_{n+1}$ 看成非基变量，在原问题的最优单纯形表中增加一列

$$\begin{bmatrix} \bar{\boldsymbol{a}}_{n+1} \\ z_{n+1}-c_{n+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^{-1}\boldsymbol{a}_{n+1} \\ \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{a}_{n+1}-c_{n+1} \end{bmatrix},$$

就得到新问题的单纯形表. 若 $z_{n+1}-c_{n+1}\leqslant 0$，则已找到新问题的最优解；否则，需继续用单纯形法求解新问题.

### 1.7.5 增加新的不等式约束

设原问题的最优单纯形表对应的约束为

$$\boldsymbol{x}_B+\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_N=\boldsymbol{B}^{-1}\boldsymbol{b}, \tag{1.75}$$

增加新约束 $\boldsymbol{a}^{(m+1)}\boldsymbol{x}\leqslant b_{m+1}$，其中，$\boldsymbol{a}^{(m+1)}$ 为 $n$ 维行向量. 用 $\boldsymbol{a}_B^{(m+1)}$ 和 $\boldsymbol{a}_N^{(m+1)}$ 分别表示 $\boldsymbol{a}^{(m+1)}$ 中与原问题的 $\boldsymbol{B}$ 和 $\boldsymbol{N}$ 对应的两部分. 引进松弛变量 $x_{n+1}$，将新约束化成等式

$$\boldsymbol{a}_B^{(m+1)}\boldsymbol{x}_B+\boldsymbol{a}_N^{(m+1)}\boldsymbol{x}_N+x_{n+1}=b_{m+1}. \tag{1.76}$$

将(1.75)代入(1.76)得

$$(a_N^{(m+1)} - a_B^{(m+1)}B^{-1}N)x_N + x_{n+1} = b_{m+1} - a_B^{(m+1)}B^{-1}b. \tag{1.77}$$

将(1.76)作为第 $m+1$ 行写在原有的最优单纯形表中，并将原有第 $m+1$ 行移作第 $m+2$ 行. 然后，通过行初等变换将第 $m+1$ 行化成(1.77)的形式，便可得到新问题的单纯形表. 若 $b_{m+1} - a_B^{(m+1)}B^{-1}b \geqslant 0$，则已找到新问题的最优解；否则，得到的是新问题的一个正则解，可由此开始用对偶单纯形法求解新问题.

**例 1.7.4** 考虑例 1.7.1 中的线性规划. 设增加约束 $-3x_1 + 2x_2 \leqslant 3$，求新问题的最优解.

**解** 引进松弛变量 $x_5$，则新约束变成

$$-3x_1 + 2x_2 + x_5 = 3.$$

将此式写到原有最优单纯形表的第 3 行，将原有第 3 行移作第 4 行，得

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | |
|---|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | −1 | 0 | 2 |
| $x_2$ | −1 | 1 | 0 | 1 | 0 | 2 |
| $x_5$ | −3 | 2 | 0 | 0 | 1 | 3 |
| $f$ | −1 | 0 | 0 | −4 | 0 | −8 |

将 $\bar{a}_{32}$ 消为 0，得

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | |
|---|---|---|---|---|---|---|
| $x_3$ | 2 | 0 | 1 | −1 | 0 | 2 |
| $x_2$ | −1 | 1 | 0 | 1 | 0 | 2 |
| $x_5$ | −1 | 0 | 0 | −2 | 1 | −1 |
| $f$ | −1 | 0 | 0 | −4 | 0 | −8 |

由于 $\bar{b}_3=-1<0$，故得到的解是新问题的一个正则解．再由此开始，用对偶单纯形法求解新问题，得最优解为 $\boldsymbol{x}=(1,3)^{\mathrm{T}}$，目标函数最优值为 $-7$．

### 1.7.6 增加新的等式约束

设增加一个新的等式约束 $\boldsymbol{a}^{(m+1)}\boldsymbol{x}=b_{m+1}$．如果原有问题的最优解满足这个新等式，则它也是新问题的最优解；否则，不妨设原有问题最优解 $\boldsymbol{x}$ 使 $\boldsymbol{a}^{(m+1)}\boldsymbol{x}<b_{m+1}$，引进人工变量 $x_{n+1}$，使新增加的约束成为等式 $\boldsymbol{a}^{(m+1)}\boldsymbol{x}+x_{n+1}=b_{m+1}$．然后，用两阶段法求解新问题．

## 1.8 Karmarkar 投影尺度算法

### 1.8.1 Karmarkar 算法的基本思想

1984 年，AT&T 贝尔实验室的 $N$ $K$ Karmarkar，提出一个新的解线性规划的多项式时间算法．这种算法不同于单纯形方法，每次迭代不是从一个极点出发求改进的极点，而是使迭代点保持在 1.8.2 节中定义的单纯形内部．因此是一种内点算法．其基本思想是，给定一个可行内点，对解空间进行变换，使得现行解位于变换空间中多胞形的中心附近，然后使它沿着最速下降方向移动，求一个改进的可行内点．再作逆变换，将在变换空间中求得的点映射回原来的解空间，得到新的内点．重复以上过程，直至求出满足精度要求的最优解．算法的总复杂性为 $O(n^{3.5}L^2\ \ln L \ln\ln L)$，其中 $n$ 是变量数，$L$ 是输入长度．

### 1.8.2 karmarkar 标准形式

Karmarkar 算法中，采用一种特殊的标准形式，后面将给出解一般线性规划转化为求解标准问题的方法，这种标准问题定义

如下：

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \ ; \\
\text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \\
& \boldsymbol{x} \in \Delta,
\end{aligned}
\tag{1.78}
$$

其中，$\boldsymbol{A}$ 是 $m\times n$ 矩阵，秩为 $m$，$\boldsymbol{c},\boldsymbol{x}\in E^n$，$\Delta=\{\boldsymbol{x}\mid \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}=1,\boldsymbol{x}\geqslant\boldsymbol{0}\}$ 称为单纯形，$\boldsymbol{e}$ 是分量全为 1 的 $n$ 维向量. 若可行解 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 的每个分量均大于零，则称 $\boldsymbol{x}$ 为内点解，并假设

(1) $\boldsymbol{A}\boldsymbol{e}=0$，即单纯形 $\Delta$ 的中心 $\frac{1}{n}\boldsymbol{e}=\left(\frac{1}{n},\cdots,\frac{1}{n}\right)^{\mathrm{T}}$ 是可行内点.

(2) 问题(1.78)的最优值是零.

下例给出一个 Karmarkar 标准问题：

**例 1.8.1**

$$
\begin{aligned}
\min \quad & x_1 + x_2 - \frac{1}{3} \ ; \\
\text{s.t.} \quad & x_1 - 2x_2 + x_3 = 0, \\
& x_1 + x_2 + x_3 = 1, \\
& x_1,\ x_2,\ x_3 \geqslant 0.
\end{aligned}
$$

显然，$\frac{1}{n}\boldsymbol{e}=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^{\mathrm{T}}$ 是一个可行内点，问题的最优解 $\boldsymbol{x}^*=\left(0,\frac{1}{3},\frac{2}{3}\right)^{\mathrm{T}}$，最优值为 0.

### 1.8.3 单纯形上的投影变换及势函数

在 Karmarkar 算法中，用到一种特殊的投影变换，其定义是

$$
T_{\bar{x}}(\boldsymbol{x}) = \frac{\boldsymbol{D}^{-1}\boldsymbol{x}}{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{D}^{-1}\boldsymbol{x}}, \quad \forall\, \boldsymbol{x} \in \Delta, \tag{1.79}
$$

其中 $\boldsymbol{D}=\mathrm{diag}(\bar{x}_1,\cdots,\bar{x}_n)$ 是对角矩阵，$\bar{\boldsymbol{x}}=(\bar{x}_1,\cdots,\bar{x}_n)^{\mathrm{T}}$ 是单纯形 $\Delta$ 的一个内点，即 $\bar{x}_i>0, i=1,\cdots,n, \sum_{i=1}^{n}\bar{x}_i=1$. 将点 $\boldsymbol{x}$ 的像记作

$x'$，即

$$x' = \frac{D^{-1}x}{e^{\mathrm{T}}D^{-1}x}.$$

映射 $T_{\bar{x}}$具有下列性质：

（1）$T_{\bar{x}}$是一一对应的映射，并将 $\Delta$ 映射成 $\Delta$. 其逆映射为

$$x = \frac{Dx'}{e^{\mathrm{T}}Dx'}. \tag{1.80}$$

（2）$T_{\bar{x}}(\bar{x})=\frac{1}{n}e$ 是 $\Delta$ 的中心.

（3）若 $x$ 是 $\Delta$ 的端点，则 $T_{\bar{x}}(x)$也是 $\Delta$ 的端点.

（4）若 $x$ 在 $\Delta$ 的边界上，则 $T_{\bar{x}}(x)$在 $\Delta$ 的边界上.

（5）若 $x$ 是 $\Delta$ 的内点，则 $T_{\bar{x}}(x)$是 $\Delta$ 的内点.

在 Karmarkar 算法中，引进了势函数的概念. 对每个线性函数 $l(x)$，定义与它相联系的**势函数**(potential function)为

$$f(x) = \sum_{j=1}^{n}\ln\frac{l(x)}{x_j} + k, \tag{1.81}$$

其中 $k$ 为常数. 势函数具有下列性质：

（1）映射(1.79)将势函数(1.81)映射成具有相同形式的函数.

（2）线性函数 $l(x)$所期望的下降量可通过势函数值的充分减小来达到.

（3）优化势函数 $f(x)$可用优化线性函数来近似.

### 1.8.4 Karmarkar 算法

考虑 Karmarkar 标准形式(1.78)，设 $\bar{x}>\mathbf{0}$ 是一个可行内点解. 投影变换(1.79)将 $x$ 映射到

$$x' = \frac{D^{-1}x}{e^{\mathrm{T}}D^{-1}x}.$$

通过公式(1.80)，用 $x$ 的像 $x'$ 来表示 $x$：

$$x = \frac{Dx'}{e^{\mathrm{T}}Dx'}. \tag{1.82}$$

按照(1.82),将 $x$ 代入 Karmarkar 标准形式(1.78),得到变换空间的相应问题

$$\left.\begin{aligned} \min \quad & \frac{c^{\mathrm{T}}Dx'}{e^{\mathrm{T}}Dx'}; \\ \text{s.t.} \quad & ADx' = \mathbf{0}, \\ & e^{\mathrm{T}}x' = 1, \\ & x' \geqslant \mathbf{0}. \end{aligned}\right\} \tag{1.83}$$

在(1.96)中,根据(1.92),$\bar{x}$ 的像 $\bar{x}' = T_{\bar{x}}(\bar{x}) = \frac{1}{n}e$ 是位于单纯形 $\Delta$ 中心的可行解.若把上述问题的约束矩阵记作

$$B = \begin{bmatrix} AD \\ e^{\mathrm{T}} \end{bmatrix} \tag{1.84}$$

则矩阵 $B$ 的零空间的任一方向 $d \in E^n$,即满足 $Bd = \mathbf{0}$,是 $\bar{x}'$ 处的可行方向.为从这些可行方向中选择一个比较好的下降方向,将(1.96)目标函数的分子 $c^{\mathrm{T}}Dx'$ 的负梯度$(-Dc)$投影到约束矩阵 $B$ 的零空间中,得到一个可行下降方向:

$$d = -[I - B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}B]Dc \tag{1.85}$$

这样,算法从原空间的一个可行解开始,通过投影变换将这个解映射到单纯形 $\Delta$ 的中心,由此出发,按(1.85)选择一个移动方向.取步长不超过单纯形 $\Delta$ 内切球的半径 $\frac{1}{\sqrt{n(n-1)}}$,移动到变换空间中一个新的内点可行解.再按(1.80)将新的解映射回原来的解空间.

计算步骤:

(1) 给定参数值 $\alpha \in (0,1]$,令 $x^{(0)} = \frac{1}{n}e$,其中 $n$ 维向量 $e = (1,1,\cdots,1)^{\mathrm{T}}$.给定足够大的终止参数 $q$,例如 $q = O(L)$,置 $k = 0$.

(2) 若$\frac{c^{\mathrm{T}}x^{(k)}}{c^{\mathrm{T}}x^{(0)}}\leqslant 2^{-q}$，则停止计算，得到最优解 $x^{(k)}$；否则，进行步骤(3).

(3) 令 $D=\mathrm{diag}[x_1^{(k)},\cdots,x_n^{(k)}]$，

$$B=\begin{bmatrix}AD\\ e^{\mathrm{T}}\end{bmatrix}.$$

(4) 令 $d_p=-[I-B^{\mathrm{T}}(BB^{\mathrm{T}})^{-1}B]Dc$.

(5) 令 $\hat{d}_p=d_p/\|d_p\|$.

(6) 计算极小点

$$b'=x^{(0)}+\frac{\alpha}{\sqrt{n(n-1)}}\hat{d}_p.$$

(7) 映射回原来的解空间，令

$$x^{(k+1)}=\frac{Db'}{e^{\mathrm{T}}Db'}.$$

(8) 计算势函数值 $f(x^{(k)})$和 $f(x^{(k+1)})$，若 $f(x^{(k)})-f(x^{(k+1)})<\delta$($\delta$ 取值参见(1.86))，则停止计算，Karmarkar 标准问题的极小值大于零；否则，进行步骤(9).

(9) 置 $k:=k+1$，返回步骤(2).

下面给出关于上述算法的两个主要定理.

**定理 1.8.2** 算法在$O(n(q+\log n))$步内求得可行点 $x$，使得$\frac{c^{\mathrm{T}}x}{c^{\mathrm{T}}x^{(0)}}\leqslant 2^{-q}$，其中 $q$ 是给定的终止参数，$x^{(0)}$是初点，取为单纯形 $\Delta$ 的中心.

运用 $f(x)=\sum_j \ln\frac{c^{\mathrm{T}}x}{x_j}$将目标函数 $c^{\mathrm{T}}x$ 与势函数 $f(x)$联系起来，有下列结论：

**定理 1.8.3** Karmarkar 算法中，对于每个 $k$，必有下列情形之一：

(1) $f(x^{(k+1)})\leqslant f(x^{(k)})-\delta$.

(2) 目标函数的极小值大于零.

其中 $\delta$ 是一个常数，依赖于步骤(1)中参数 $\alpha$ 的选择. 例如，若 $\alpha=1/4$，则 $\delta=1/8$. $\alpha$ 与 $\delta$ 有下列关系：

$$\delta = \alpha - \frac{\alpha^2}{2} - \frac{\alpha^2 n}{(n-1)[1-\alpha\sqrt{n/(n-1)}]}. \quad (1.86)$$

根据该定理，在运用 Karmarkar 算法的过程中，若出现情形(2)，则停止计算，得出目标函数值大于零的结论，则原来的问题不存在有限最优值，即或者是不可行的，或者是无界的.

### 1.8.5 向 Karmarkar 标准形式的转换

考虑一般线性规划问题：

$$\begin{aligned} \min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \ ; \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned} \quad (1.87)$$

可用多种方法将(1.87)的求解转化为解 Karmarkar 标准问题. 下面介绍 Karmarkar 运用对偶理论实现这种转换的一种方法.

(1.87)的对偶问题：

$$\begin{aligned} \max \quad & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} \ ; \\ \text{s.t.} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} \leqslant \boldsymbol{c}, \boldsymbol{u} \geqslant \boldsymbol{0}. \end{aligned} \quad (1.88)$$

组合(1.87)与(1.88)，得到

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} & \geqslant \boldsymbol{b}, \\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} & \leqslant \boldsymbol{c}, \\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} & = 0, \\ \boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{u} & \geqslant \boldsymbol{0}. \end{aligned} \quad (1.89)$$

根据对偶理论，(1.89)可行的充要条件是(1.87)存在有限最优解. 引进松弛变量 $\boldsymbol{y}$ 和 $\boldsymbol{v}$，将(1.89)表示成

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{b},$$

$$A^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{c},$$
$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} = 0, \tag{1.90}$$
$$\boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{u} \geqslant \boldsymbol{0}, \boldsymbol{y} \geqslant \boldsymbol{0}, \boldsymbol{v} \geqslant \boldsymbol{0}.$$

设 $\boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{u}_0, \boldsymbol{v}_0$ 是在正卦限内的严格内点，引进人工变量 $\lambda$，求解下列问题：

$$\begin{aligned}
\min \quad & \lambda; \\
\text{s.t.} \quad & A\boldsymbol{x} - \boldsymbol{y} + (\boldsymbol{b} - A\boldsymbol{x}_0 + \boldsymbol{y}_0)\lambda = \boldsymbol{b}, \\
& A^{\mathrm{T}}\boldsymbol{u} + \boldsymbol{v} + (\boldsymbol{c} - A^{\mathrm{T}}\boldsymbol{u}_0 - \boldsymbol{v}_0)\lambda = \boldsymbol{c}, \\
& \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u} + (-\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}_0 + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{u}_0)\lambda = \boldsymbol{0}, \\
& \boldsymbol{x} \geqslant \boldsymbol{0},\ \boldsymbol{y} \geqslant \boldsymbol{0},\ \boldsymbol{u} \geqslant \boldsymbol{0},\ \boldsymbol{v} \geqslant \boldsymbol{0},\ \lambda \geqslant \boldsymbol{0}.
\end{aligned} \tag{1.91}$$

显然，$\boldsymbol{x}=\boldsymbol{x}_0, \boldsymbol{y}=\boldsymbol{y}_0, \boldsymbol{u}=\boldsymbol{u}_0, \boldsymbol{v}=\boldsymbol{v}_0, \lambda=1$ 是一个严格内点可行解. $\lambda$ 极小值为 0 的充要条件是(1.90)有可行解.

(1.91)是变量 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}, \lambda$ 的线性规划，为表达方便，不妨用一个新的 $\boldsymbol{x}$ 表示所有这些变量，则(1.91)可以改写成

$$\begin{aligned}
\min \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}; \\
\text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b}, \\
& \boldsymbol{x} \geqslant \boldsymbol{0}.
\end{aligned} \tag{1.92}$$

值得注意，这里的 $A, \boldsymbol{c}, \boldsymbol{b}, \boldsymbol{x}$ 与(1.91)中同样符号含意不同. 不失一般性，假设 $\boldsymbol{c}, \boldsymbol{x} \in E^n$，$A$ 为 $m \times n$ 矩阵. 为把(1.92)转化为 Karmarkar 标准形式，将 $E^n$ 中的正卦限映射变换成一个单纯形. 定义变换

$$\boldsymbol{x}' = T(\boldsymbol{x}),\ \boldsymbol{x} \in E^n,\ \boldsymbol{x}' \in E^{n+1}, \tag{1.93}$$

具体形式是

$$\begin{aligned}
x_i' &= \frac{x_i/a_i}{\sum_j \frac{x_j}{a_j} + 1},\ i = 1, \cdots, n, \\
x_{n+1}' &= 1 - \sum_{i=1}^{n} x_i',
\end{aligned} \tag{1.94}$$

其中 $\boldsymbol{a}=(a_1,\cdots,a_n)^{\mathrm{T}}$ 为已知的内部起始点，每个分量均大于零. $T(\boldsymbol{x})$的逆变换是

$$x_i=\frac{a_i x_i'}{x_{n+1}'},\quad i=1,\cdots,n. \tag{1.95}$$

变换(1.94)将 $E^n$ 中的正卦限

$$P_+=\{\boldsymbol{x}\in E^n \mid \boldsymbol{x}\geqslant \boldsymbol{0}\} \tag{1.96}$$

映射成 $E^{n+1}$中的单纯形

$$\Delta=\{\boldsymbol{x}\in E^{n+1} \mid \boldsymbol{x}\geqslant 0,\sum_{i=1}^{n+1}x_i=1\}. \tag{1.97}$$

将(1.92)中 $\boldsymbol{A}$ 的第 $i$ 列记作 $\boldsymbol{A}_i$，定义矩阵 $\boldsymbol{A}'$ 为

$$\begin{aligned} \boldsymbol{A}_i' &= a_i\boldsymbol{A}_i,\ i=1,\cdots,n, \\ \boldsymbol{A}_{n+1}' &= -\boldsymbol{b}, \end{aligned} \tag{1.98}$$

由(1.95)，(1.98)及 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$，易知

$$\sum_{i=1}^{n+1}\boldsymbol{A}_i'x_i'=\boldsymbol{0},\text{即 } \boldsymbol{A}'\boldsymbol{x}'=\boldsymbol{0}. \tag{1.99}$$

定义 $\boldsymbol{c}'\in E^{n+1}$如下：

$$\begin{aligned} c_i' &= a_i c_i,\quad i=1,\cdots,n, \\ c_{n+1}' &= 0, \end{aligned}$$

则由 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}=0$ 推出 $\boldsymbol{c}'^{\mathrm{T}}\boldsymbol{x}'=0$. 得到变换问题

$$\begin{aligned} \min\quad & \boldsymbol{c}'^{\mathrm{T}}\boldsymbol{x}'\qquad ; \\ \text{s.t.}\quad & \boldsymbol{A}'\boldsymbol{x}'=\boldsymbol{0}, \\ & \sum_{i=1}^{n+1}x_i'=1, \\ & \boldsymbol{x}'\geqslant \boldsymbol{0}. \end{aligned} \tag{1.100}$$

(1.100)属于 Karmarkar 标准形式，可用 1.8.4 节提供的算法进行求解.

# 参 考 文 献

1 张建中,许绍吉.线性规划.科学出版社,1990

2 管梅谷,郑汉鼎.线性规划.山东科学技术出版社,1983

3 赵凤治.线性规划计算方法.科学出版社,1981

4 陈宝林.最优化理论与算法.清华大学出版社,1989

5 Bazaraa M S and Jarvis J J. Linear programming and Network Flows. New York: Wiley, 1977

6 Luenberger D G. Introduction to Linear and Nonlinear programming. Addison-Wesley, 1973(中译本.夏尊铨等译.线性与非线性规划引论.科学出版社,1982)

7 Karmarkar N. A new polynomial time algorithm for linear programming. Combinatorica 4, 1984, 373～395

# 2 非线性规划

## 2.1 基础知识

### 2.1.1 非线性规划问题

非线性规划处理的问题是在等式和/或不等式约束下优化某个目标函数,求出最优解. 一般表示成:

$$\left.\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \geqslant 0,\ i = 1,\cdots,m, \\ & h_j(\boldsymbol{x}) = 0,\ j = 1,\cdots,l. \end{aligned}\right\} \tag{2.1}$$

其中 $\boldsymbol{x} \in E^n$,$f(\boldsymbol{x})$为**目标函数**(objective function),$g_i(\boldsymbol{x})$,$h_j(\boldsymbol{x})$为**约束函数**(constraint function),这些函数中,至少有一个是非线性函数.

约束条件有时写成集约束形式,令 $S=\{\boldsymbol{x} \mid g_i(\boldsymbol{x})\geqslant 0, i=1,\cdots, m; h_j(\boldsymbol{x})=0, j=1,\cdots,l\}$,称 $S$ 为**可行集**或**可行域**(feasible region). $S$ 中的点称为**可行点**. 这样,(2.1)中的约束条件可用集约束表示.

$$\left.\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & \boldsymbol{x} \in S. \end{aligned}\right\} \tag{2.2}$$

特别,当 $E^n$ 中每一点均为可行点时,(2.2)即

$$\min \quad f(\boldsymbol{x}), \boldsymbol{x} \in E^n. \tag{2.3}$$

(2.3)称为**无约束最优化问题**(unconstrained optimization problem).

**例 2.1.1** 设有 $n$ 个市场,第 $j$ 个市场的位置为$(a_j,b_j)$,对某

种货物的需要量为 $q_j$，$j=1,\cdots,n$. 现计划建立 $m$ 个货栈，第 $i$ 个货栈的容量为 $c_i$，$i=1,\cdots,m$. 试确定货栈的位置，使各货栈到各市场的运输量与路程乘积之和最小.

建立数学模型. 设第 $i$ 个货栈的位置为$(x_i,y_i)$，$i=1,\cdots,m$. 第 $i$ 个货栈供给第 $j$ 个市场的货物量为 $w_{ij}$，$i=1,\cdots,m$，$j=1,\cdots,n$. 第 $i$ 个货栈到第 $j$ 个市场的距离为 $d_{ij}$，定义为

$$d_{ij}=\sqrt{(x_i-a_j)^2+(y_i-b_j)^2},\tag{2.4}$$

或

$$d_{ij}=|x_i-a_j|+|y_i-b_j|.\tag{2.5}$$

目标是使运输量与路程乘积之和为最小，如果距离按式(2.4)定义，就是使

$$\sum_{i=1}^{m}\sum_{j=1}^{n}w_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2}$$

最小. 约束条件是

(1) 每个货栈向各市场提供的货物量之和不能超过它的容量；

(2) 每个市场从各货栈得到的货物量之和应等于它的需要量；

(3) 运输量不能为负数.

数学模型如下：

$$\begin{aligned}
\min\quad & \sum_{i=1}^{m}\sum_{j=1}^{n}w_{ij}\sqrt{(x_i-a_j)^2+(y_i-b_j)^2};\\
\text{s.t.}\quad & \sum_{j=1}^{n}w_{ij}\leqslant c_i\quad i=1,\cdots,m,\\
& \sum_{i=1}^{m}w_{ij}=q_j\quad j=1,\cdots,n,\\
& w_{ij}\geqslant 0\quad i=1,\cdots,m;j=1,\cdots,n.
\end{aligned}$$

由于目标函数是非线性函数，因此是非线性规划.

下面给出最优解概念.

**定义 2.1.2** 设 $f(x)$为目标函数，$S$ 为可行域，$x^* \in S$，若对每一个 $x\in S$，均成立 $f(x)\geq f(x^*)$，则称 $x^*$ 为如下极小化问题的**最优解**（**整体最优解**）(global optimal solution).

$$\begin{aligned} &\min \quad f(x); \\ &\text{s.t.} \quad x \in S. \end{aligned}$$

**定义 2.1.3** 设 $f(x)$为目标函数，$S$ 为可行域，$x^* \in S$，若存在 $x^*$ 的 $\varepsilon$ 邻域

$$N_\varepsilon(x^*) = \{x \mid \|x - x^*\| < \varepsilon, \varepsilon > 0\},$$

使得对每个 $x\in S\cap N_\varepsilon(x^*)$，成立 $f(x)\geq f(x^*)$，则称 $x^*$ 为极小化问题

$$\begin{aligned} &\min \quad f(x); \\ &\text{s.t.} \quad x \in S \end{aligned}$$

的**局部最优解**(local optimal solution).

对于极大化问题，可类似地定义整体最优解和局部最优解.

## 2.1.2 凸集与凸函数

**凸集**(convex set)和**凸函数**(convex function)在非线性规划的理论中具有重要作用. 下面给出凸集和凸函数的一些基本知识.

### 1 凸集

**定义 2.1.4** 设 $S$ 为 $E^n$ 中一个集合，若对 $S$ 中任意两点，连结它们的线段仍属于 $S$；换言之，对 $S$ 中任意两点 $x^{(1)}, x^{(2)}$ 及每个实数 $\lambda\in[0,1]$，都有

$$\lambda x^{(1)} + (1-\lambda)x^{(2)} \in S,$$

则称 $S$ 为**凸集**. $\lambda x^{(1)}+(1-\lambda)x^{(2)}$称为 $x^{(1)}$和 $x^{(2)}$的**凸组合**(convex combination).

**例 2.1.5** **超平面**(hyperplane)$H \triangleq \{x \mid p^{\mathrm{T}}x=\alpha\}$为凸集.

**例 2.1.6** 半空间(half space)$H^- \triangleq \{\boldsymbol{x} \mid \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \leqslant \alpha\}$为凸集.

**例 2.1.7** 射线 $L=\{\boldsymbol{x} \mid \boldsymbol{x}=\boldsymbol{x}^{(0)}+\lambda \boldsymbol{d}, \lambda \geqslant 0\}$为凸集. 其中 $\boldsymbol{d}$ 是给定的非零向量,$\boldsymbol{x}^{(0)}$为定点.

在凸集中,比较重要的特殊情形有凸锥和多面集.

**定义 2.1.8** 设有集合 $C \subset E^n$,若对 $C$ 中每一点 $x$,当 $\lambda$ 取任何非负数时都有 $\lambda x \in C$,则称 $C$ 为**锥**,若 $C$ 又为凸集,则称 $C$ 为**凸锥**(convex cone).

**例 2.1.9** $n$ 维向量 $\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \cdots, \boldsymbol{\alpha}^{(k)}$的所有非负线性组合

$$\left\{\sum_{i=1}^{k} \lambda_i \boldsymbol{\alpha}^{(i)} \mid \lambda_i \geqslant 0, i=1, \cdots, k\right\}$$

为凸锥.

**定义 2.1.10** 有限个半空间的交$\{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\}$称为**多面集**(polyhedral set). 其中 $\boldsymbol{A}$ 为 $m \times n$ 矩阵,$\boldsymbol{b}$ 为 $m$ 维向量. 在凸集的理论及应用中,极点和极方向的概念有着重要作用.

**定义 2.1.11** 设 $S$ 为非空凸集,$\boldsymbol{x} \in S$,若 $\boldsymbol{x}$ 不能表示成 $S$ 中两个不同点的凸组合;换言之,若假设

$$\boldsymbol{x}=\lambda \boldsymbol{x}^{(1)}+(1-\lambda) \boldsymbol{x}^{(2)}, \lambda \in(0,1),$$

$\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)} \in S$,必推得 $\boldsymbol{x}=\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(2)}$,则称 $\boldsymbol{x}$ 是凸集 $S$ 的**极点**(extreme point).

**定义 2.1.12** 设 $S$ 为 $E^n$ 中的闭凸集,$\boldsymbol{d}$ 为非零向量,如果对 $S$ 中的每一个 $\boldsymbol{x}$,都有射线

$$\{\boldsymbol{x}+\lambda \boldsymbol{d} \mid \lambda \geqslant 0\} \subset S,$$

则称向量 $\boldsymbol{d}$ 为 $S$ 的**方向**. 又设 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}$是 $S$ 的两个方向,若对任何正数 $\lambda$,都有 $\boldsymbol{d}^{(1)} \neq \lambda \boldsymbol{d}^{(2)}$,则称 $\boldsymbol{d}^{(1)}$和 $\boldsymbol{d}^{(2)}$是两个不同的方向. 若 $S$ 的方向 $\boldsymbol{d}$ 不能表示成该集合的两个不同方向的正的线性组合,则称 $\boldsymbol{d}$ 为 $S$ 的**极方向**(extreme direction).

显然,有界集不存在方向,因而也不存在极方向. 对于无界集才有方向的概念. 下面给出多面集的一个重要性质.

**定理 2.1.13** 表示定理(representation theorem) 设 $S=\{x \mid Ax=b, x\geq 0\}$为非空多面集,则有

(1) 极点集非空,且存在有限个极点. $x^{(1)},\cdots,x^{(k)}$.

(2) 极方向集合为空集的充要条件是 $S$ 有界. 若 $S$ 无界,则存在有限个极方向. $d^{(1)},\cdots,d^{(l)}$.

(3) $x\in S$ 的充要条件是

$$
\begin{aligned}
x &= \sum_{j=1}^{k}\lambda_j x^{(j)} + \sum_{j=1}^{l}\mu_j d^{(j)}, \\
&\sum_{j=1}^{k}\lambda_j = 1, \\
&\lambda_j \geq 0 \quad j=1,\cdots,k, \\
&\mu_j \geq 0 \quad j=1,\cdots,l.
\end{aligned}
$$

根据表示定理,对于多面集只需研究其极点和极方向.

下面是一般凸集所具有的重要性质.

**定理 2.1.14** 设 $S$ 是 $E^n$ 中的非空闭凸集,$y\notin S$,则存在非零向量 $p$ 及数 $\varepsilon>0$,使得对每个 $x\in S$,成立 $p^{\mathrm{T}}y\geq\varepsilon+p^{\mathrm{T}}x$.

定理表明,当 $S$ 为闭凸集,$y\notin S$ 时,能够做一个以 $p$ 为法向量的超平面,使点 $y$ 和凸集 $S$ 各在一边,即将 $y$ 和 $S$ 分离. 利用点和凸集分离定理,易证在凸集每一边界点处存在**支撑超平面**,使凸集位于此超平面的一侧.

**定理 2.1.15** 设 $S$ 是 $E^n$ 中一个非空凸集,$y\in\partial S$,则存在非零向量 $p$,使得对每一 $x\in \mathrm{cl}S$,成立 $p^{\mathrm{T}}y\geq p^{\mathrm{T}}x$.

其中 $\partial S$ 表示集 $S$ 的边界;$\mathrm{cl}S$ 表示集合 $S$ 的**闭包**(closure),即 $S$ 的内点与边界点组成的集合.

若两个凸集的交集为空集,则可用超平面将其分离.

**定理 2.1.16** 设 $S_1$ 和 $S_2$ 是 $E^n$ 中两个非空凸集,且 $S_1\cap S_2=\varnothing$,则存在非零向量 $p$,使

$$\inf\{\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \mid \boldsymbol{x} \in S_1\} \geqslant \sup\{\boldsymbol{p}^{\mathrm{T}}\boldsymbol{x} \mid \boldsymbol{x} \in S_2\}.$$

利用凸集分离定理，可以证明 Farkas 定理和 Gordan 定理，它们在最优化理论的研究中具有重要作用.

**定理 2.1.17**(Farkas 定理)　设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，$\boldsymbol{c}$ 为 $n$ 维向量，则 $\boldsymbol{A}\boldsymbol{x}\leqslant\boldsymbol{0}, \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}>0$ 有解的充要条件是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{c}, \boldsymbol{y}\geqslant\boldsymbol{0}$ 无解.

这个定理把不等式组是否有解与方程组是否无非负解相联系.

**定理 2.1.18**(Gordan 定理)　设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵，则 $\boldsymbol{A}\boldsymbol{x}<\boldsymbol{0}$ 有解的充要条件是不存在非零向量 $\boldsymbol{y}\geqslant\boldsymbol{0}$，使 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=0$

**2　凸函数**

**定义 2.1.19**　设 $S$ 为 $E^n$ 中的非空凸集，$f$ 是定义在 $S$ 上的实函数. 如果对任意的 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\in S$ 及每个数 $\lambda\in(0,1)$，都有

$$f(\lambda\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) \leqslant \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda)f(\boldsymbol{x}^{(2)}),$$

则称 $f$ 为 $S$ 上的**凸函数**(convex function).

如果对任意相异的 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\in S$，及每个数 $\lambda\in(0,1)$，都有

$$f(\lambda\boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}) < \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda)f(\boldsymbol{x}^{(2)}),$$

则称 $f$ 为 $S$ 上的**严格凸函数**(strictly convex function).

如果 $-f$ 为 $S$ 上的凸函数，则称 $f$ 为 $S$ 上的**凹函数**(concave function).

下面列举凸函数的一些性质.

**定理 2.1.20**　设 $f$ 是定义在凸集 $S$ 上的凸函数，实数 $\lambda\geqslant0$，则 $\lambda f$ 也是定义在 $S$ 上的凸函数.

**定理 2.1.21**　设 $f_1$ 和 $f_2$ 是定义在凸集 $S$ 上的凸函数，则 $f_1+f_2$ 也是定义在 $S$ 上的凸函数.

由上述两个性质可知，有限个凸函数的非负线性组合仍为凸

函数.

**定理 2.1.22** 设 $S$ 是 $E^n$ 中一个非空凸集，$f$ 是定义在 $S$ 上的凸函数，则**水平集**(level set)$S_\alpha=\{x \mid x\in S, f(x)\leqslant\alpha\}$是凸集.

关于连续性，有下列结论：

**定理 2.1.23** 设 $S$ 是 $E^n$ 中一个凸集，$f$ 是定义在 $S$ 上的凸函数，则 $f$ 在 $S$ 的内部连续.

**定理 2.1.24** 设 $f$ 是一个凸函数，$\boldsymbol{x}\in E^n$，在 $\boldsymbol{x}$ 处 $f(\boldsymbol{x})$取有限值，则 $f$ 在 $\boldsymbol{x}$ 处沿任何方向 $\boldsymbol{d}$ 的右侧导数及左侧导数都存在(包括$\pm\infty$).

凸函数的重要性在于下面的基本性质.

**定理 2.1.25** 设 $S$ 是 $E^n$ 中非空凸集，$f$ 是定义在 $S$ 上的凸函数，则 $f$ 在 $S$ 上的局部极小点是整体极小点，且极小点的集合为凸集.

运用凸函数的定义及性质判断一个函数是否为凸函数，一般说来，比较复杂.下面给出可微函数为凸函数的充分必要条件，运用这些条件不难判断一个可微函数是否为凸函数.先给出梯度和 Hesse 矩阵的定义.

**定义 2.1.26** 设函数 $f(\boldsymbol{x})$存在一阶偏导数，$\boldsymbol{x}\in E^n$，则称向量

$$\nabla f(\boldsymbol{x})=\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}\quad\frac{\partial f(\boldsymbol{x})}{\partial x_2}\quad\cdots\quad\frac{\partial f(\boldsymbol{x})}{\partial x_n}\right)^{\mathrm{T}}$$

为 $f(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处的**梯度**(gradient).

梯度的方向是在该点函数值上升最快的方向.

**定义 2.1.27** 设函数 $f(\boldsymbol{x})$存在二阶偏导数，$\boldsymbol{x}\in E^n$，则称矩阵

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

为 $f(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处的 **Hesse 矩阵**.

**例 2.1.28** 设二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c,$$

其中 $\boldsymbol{A}$ 为 $n$ 阶对称方阵,$\boldsymbol{x}$,$\boldsymbol{b}$ 为 $n$ 维列向量,$c$ 为常数.则 $f(\boldsymbol{x})$在 $\boldsymbol{x}$ 处的梯度及 Hesse 矩阵分别为

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b},$$
$$\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}.$$

下面列举可微凸函数的一阶、二阶判别条件:

**定理 2.1.29** 设 $S$ 是 $E^n$ 中非空开凸集,$f(\boldsymbol{x})$是定义在 $S$ 上的可微函数,则 $f(\boldsymbol{x})$为凸函数的充要条件是对任意两点 $\boldsymbol{x}^{(1)}$,$\boldsymbol{x}^{(2)} \in S$,成立

$$f(\boldsymbol{x}^{(2)}) \geqslant f(\boldsymbol{x}^{(1)}) + \nabla f(\boldsymbol{x}^{(1)})^{\mathrm{T}}(\boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}).$$

而 $f(\boldsymbol{x})$为严格凸函数的充要条件是对任意相异的 $\boldsymbol{x}^{(1)}$,$\boldsymbol{x}^{(2)} \in S$,成立

$$f(\boldsymbol{x}^{(2)}) > f(\boldsymbol{x}^{(1)}) + \nabla f(\boldsymbol{x}^{(1)})^{\mathrm{T}}(\boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}).$$

此条件称为可微函数的一阶判别条件.

**定理 2.1.30** 设 $S$ 是 $E^n$ 中非空开凸集,$f(\boldsymbol{x})$是定义在 $S$ 上的二次可微函数,则 $f(\boldsymbol{x})$是凸函数的充要条件为在每一点 $\boldsymbol{x} \in S$ 处 Hesse 矩阵是半正定的.

**定理 2.1.31** 设 $S$ 是 $E^n$ 中非空开凸集,$f(\boldsymbol{x})$是定义在 $S$ 上的二次可微函数,如果在每一点 $\boldsymbol{x} \in S$,Hesse 矩阵是正定的,则

$f(\boldsymbol{x})$为严格凸函数.

**注意** 最后一个定理的逆定理不成立.若 $f(\boldsymbol{x})$是定义在 $S$ 上的严格凸函数,则在每一点 $\boldsymbol{x}\in S$ 处,Hesse 矩阵是半正定的.

利用以上几个定理容易判别一个可微函数是否为凸函数,特别对于二次函数,是很方便的.

**例 2.1.32** 给定二次函数

$$
\begin{aligned}
f(x_1,x_2) &= 2x_1^2 + x_2^2 - 2x_1x_2 + x_1 + 1 \\
&= \frac{1}{2}(x_1,x_2)\begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + x_1 + 1.
\end{aligned}
$$

由于在每一点$(x_1,x_2)$处 Hesse 矩阵

$$
\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}
$$

是正定的,因此 $f(\boldsymbol{x})$是严格凸函数.

### 3 凸规划

**定义 2.1.33** 若在式(2.1)中,$f(\boldsymbol{x})$是凸函数,$g_i(\boldsymbol{x})$是凹函数,$h_j(\boldsymbol{x})$是线性函数,则问题为求凸函数在凸集上的极小点,这类问题称为**凸规划**(convex program).

凸规划是非线性规划中的一种重要特殊情形,它具有很好的性质,正如定理 2.1.25 所述,凸规划的局部极小点就是整体极小点,且极小点集合是凸集.如果凸规划的目标函数是严格凸函数,又存在极小点,则它的极小点是唯一的.

## 2.1.3 无约束问题的极值条件

考虑问题

$$
\min \quad f(\boldsymbol{x}), \boldsymbol{x} \in E^n.
$$

先给出**局部极小点**(local minimum point)的一阶必要条件.

**定理 2.1.34** 设函数 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处可微,若 $\bar{\boldsymbol{x}}$ 是局部极小点,则梯度$\nabla f(\bar{\boldsymbol{x}})=0$.

再利用 $f(\boldsymbol{x})$的 Hesse 矩阵给定局部极小点的二阶必要条件.

**定理 2.1.35** 设函数 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处二次可微，若 $\bar{\boldsymbol{x}}$ 是局部极小点，则梯度$\nabla f(\bar{\boldsymbol{x}})=0$，并且 Hesse 矩阵$\nabla^2 f(\bar{\boldsymbol{x}})$是半正定的.

上述条件均非充分条件，阐述充分条件需要 Hesse 矩阵具有正定性.

**定理 2.1.36** 设函数 $f(\boldsymbol{x})$在点 $\bar{\boldsymbol{x}}$ 处二次可微，若梯度$\nabla f(\bar{\boldsymbol{x}})=0$，且 Hesse 矩阵$\nabla^2 f(\bar{\boldsymbol{x}})$正定，则 $\bar{\boldsymbol{x}}$ 是局部极小点.

**例 2.1.37** 利用极值条件求解下列问题：

$$\min \quad f(\boldsymbol{x}) \triangleq \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_2^2 - x_1.$$

**解** 根据 $f(\boldsymbol{x})$的定义，有

$$\frac{\partial f}{\partial x_1} = x_1^2 - 1, \frac{\partial f}{\partial x_2} = x_2^2 - 2x_2.$$

令$\nabla f(\boldsymbol{x})=0$，即 $x_1^2-1=0, x_2^2-2x_2=0$ 解此方程组得到驻点

$$\boldsymbol{x}^{(1)} = \begin{bmatrix}1\\0\end{bmatrix}, \boldsymbol{x}^{(2)} = \begin{bmatrix}1\\2\end{bmatrix}, \boldsymbol{x}^{(3)} = \begin{bmatrix}-1\\0\end{bmatrix}, \boldsymbol{x}^{(4)} = \begin{bmatrix}-1\\2\end{bmatrix}.$$

函数 $f(\boldsymbol{x})$的 Hesse 矩阵为

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix}2x_1 & 0\\0 & 2x_2-2\end{bmatrix}.$$

由此可知，在点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \boldsymbol{x}^{(4)}$处的 Hesse 矩阵依次是

$$\nabla^2 f(\boldsymbol{x}^{(1)}) = \begin{bmatrix}2 & 0\\0 & -2\end{bmatrix}, \nabla^2 f(\boldsymbol{x}^{(2)}) = \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix},$$

$$\nabla^2 f(\boldsymbol{x}^{(3)}) = \begin{bmatrix}-2 & 0\\0 & -2\end{bmatrix}, \nabla^2 f(\boldsymbol{x}^{(4)}) = \begin{bmatrix}-2 & 0\\0 & 2\end{bmatrix}.$$

由于矩阵$\nabla^2 f(\boldsymbol{x}^{(1)})$，$\nabla^2 f(\boldsymbol{x}^{(4)})$不定，根据定理 2.1.35，$\boldsymbol{x}^{(1)}$和 $\boldsymbol{x}^{(4)}$不是极小点. 在点 $\boldsymbol{x}^{(3)}$矩阵$\nabla^2 f(\boldsymbol{x}^{(3)})$是负定的，因此 $\boldsymbol{x}^{(3)}$也不是极小点，实际上它是极大点. 矩阵$\nabla^2 f(\boldsymbol{x}^{(2)})$正定，根据定理 2.1.36，$\boldsymbol{x}^{(2)}$是局部极小点.

### 2.1.4 约束问题的最优性条件

考虑约束问题：

$$\left.\begin{array}{ll} \min & f(\boldsymbol{x}); \\ \text{s.t.} & g_i(\boldsymbol{x}) \geqslant 0, \quad i=1,\cdots,m, \\ & h_j(\boldsymbol{x}) = 0, \quad j=1,\cdots,l. \end{array}\right\} \tag{2.6}$$

可行域为

$$S=\{\boldsymbol{x} \mid g_i(\boldsymbol{x}) \geqslant 0, i=1,\cdots,m; h_j(\boldsymbol{x})=0, j=1,\cdots,l\}.$$

下面介绍几个经常用到的概念.

**定义 2.1.38** 设 $f(\boldsymbol{x})$是 $E^n$ 上的实函数，$\boldsymbol{d}$ 是非零向量，$\bar{\boldsymbol{x}} \in E^n$. 若存在数 $\delta>0$，使得对每个实数 $\lambda \in (0,\delta)$，都有 $f(\bar{\boldsymbol{x}}+\lambda \boldsymbol{d})<f(\bar{\boldsymbol{x}})$，则称 $\boldsymbol{d}$ 为函数 $f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处的**下降方向**(descent direction).

如果 $f(\boldsymbol{x})$是可微函数，且$\nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{d}<0$，则 $\boldsymbol{d}$ 必为 $f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处的下降方向.

**定义 2.1.39** 设集合 $S \subset E^n$，$\bar{\boldsymbol{x}} \in \operatorname{cl} S$，$\boldsymbol{d}$ 是非零向量，若存在数 $\delta>0$，使得对每一个数 $\lambda \in (0,\delta)$，都有 $\bar{\boldsymbol{x}}+\lambda \boldsymbol{d} \in S$，则称 $\boldsymbol{d}$ 为集合 $S$ 在 $\bar{\boldsymbol{x}}$ 处的**可行方向**(feasible direction).

$S$ 在 $\bar{\boldsymbol{x}}$ 处所有可行方向组成的集合 $D=\{\boldsymbol{d} \mid \boldsymbol{d} \neq \boldsymbol{0}, \bar{\boldsymbol{x}} \in \operatorname{cl} S, \exists\, \delta>0$，使得$\forall\, \lambda \in (0,\delta)$，有 $\bar{\boldsymbol{x}}+\lambda \boldsymbol{d} \in S\}$称为在 $\bar{\boldsymbol{x}}$ 处的**可行方向锥**(cone of feasible directions).

**定义 2.1.40** 设 $S$ 是 $E^n$ 中一个非空集合，$\bar{\boldsymbol{x}} \in \operatorname{cl} S$，集合 $T=\{\boldsymbol{d} \mid$ 存在 $\boldsymbol{x}^{(k)} \in S$，$\boldsymbol{x}^{(k)} \to \bar{\boldsymbol{x}}$，以及 $\lambda_k>0$，使得 $\boldsymbol{d}=\lim\limits_{k \to \infty} \lambda_k(\boldsymbol{x}^{(k)}-\bar{\boldsymbol{x}})\}$，则称 $T$ 为集合 $S$ 在点 $\bar{\boldsymbol{x}}$ 的**切锥**(tangent cone).

根据上述定义，如果序列$\{\boldsymbol{x}^{(k)}\} \subset S$ 收敛于 $\bar{\boldsymbol{x}}$，$\boldsymbol{x}^{(k)} \neq \bar{\boldsymbol{x}}$，使得

$$\lim_{k \to \infty} (\boldsymbol{x}^{(k)}-\bar{\boldsymbol{x}}) / \|\boldsymbol{x}^{(k)}-\bar{\boldsymbol{x}}\| = \boldsymbol{d},$$

则方向 $\boldsymbol{d} \in T$.

**定义 2.1.41** 问题(2.6)中，任给一点 $\bar{\boldsymbol{x}} \in S$ 后，有些不等式

约束在 $\bar{x}$ 处成立等式，它们的下标集不妨用 $I$ 表示，即$\forall\ i\in I$，均成立 $g_i(\bar{x})=0$，通常将约束条件 $g_i(x)\geqslant 0(i\in I)$和等式约束 $h_j(x)=0(j=1,\cdots,l)$一并称为在 $\bar{x}$ 处**起作用约束**(active constraint).

起作用约束在 $\bar{x}$ 的邻域限制了可行点的范围，也就是说，当点沿某些方向稍微离开 $\bar{x}$ 时，仍能满足这些约束条件；而沿着另一些方向离开 $\bar{x}$ 时，不论步长多么小，都将违背这些约束条件.

其余约束情形则不同，当点稍微离开 $\bar{x}$ 时，不论沿什么方向，都不会违背，这些约束称为在 $\bar{x}$ 处**不起作用约束**.

这里先给出一阶必要条件：

**定理 2.1.42**(Fritz John 条件)　设在问题(2.6)中，$\bar{x}$ 为可行点，$I=\{i\,|\,g_i(\bar{x})=0\}$，$f$ 和 $g_i(i\in I)$在 $\bar{x}$ 可微，$g_i(i\notin I)$在点 $\bar{x}$ 连续，$h_j(j=1,\cdots,l)$在 $\bar{x}$ 连续可微. 如果 $\bar{x}$ 是局部最优解，则存在不全为零的数 $w_0,w_i(i\in I)$和 $v_j(j=1,\cdots,l)$，使得

$$w_0\nabla f(\bar{x})-\sum_{i\in I}w_i\nabla g_i(\bar{x})-\sum_{j=1}^{l}v_j\nabla h_j(\bar{x})=0,$$
$$w_0,w_i\geqslant 0,i\in I.$$

通常将满足 Fritz John 条件的点称为 Fritz John 点.

**定理 2.1.43**(Kuhn-Tucker 定理)　设在问题(2.6)中，$\bar{x}$ 为可行点，$I=\{i\,|\,g_i(\bar{x})=0\}$，$f$ 和 $g_i(i\in I)$在点 $\bar{x}$ 可微，$g_i(i\notin I)$在点 $\bar{x}$ 连续，$h_j(j=1,\cdots,l)$在点 $\bar{x}$ 连续可微，向量集$\{\nabla g_i(\bar{x}),\nabla h_j(\bar{x})\,|\,i\in I,j=1,\cdots,l\}$线性无关. 如果 $\bar{x}$ 是局部最优解，则存在非负数 $w_i(i\in I)$和数 $v_j(j=1,\cdots,l)$，使得

$$\nabla f(\bar{x})-\sum_{i\in I}w_i\nabla g_i(\bar{x})-\sum_{j=1}^{l}v_j\nabla h_j(\bar{x})=0.$$

由上述定理知，在最优解处，目标函数的梯度可用起作用约束梯度的非负线性组合及等式约束梯度的线性组合来表示. 上述条件还可写成等价形式：

$$\nabla f(\bar{x}) - \sum_{i=1}^{m} w_i \nabla g_i(\bar{x}) - \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) = 0. \quad (2.7)$$

$$w_i g_i(\bar{x}) = 0, i = 1, \cdots, m. \quad (2.8)$$

(2.8)称为**互补松弛条件**(complementary slackness condition).

下面定义 **Lagrange 函数**.

$$L(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}) = f(\boldsymbol{x}) - \sum_{i=1}^{m} w_i g_i(\boldsymbol{x}) - \sum_{j=1}^{l} v_j h_j(\boldsymbol{x}) \quad (2.9)$$

(2.7)可写作$\nabla_x L(\bar{\boldsymbol{x}}, \boldsymbol{w}, \boldsymbol{v}) = 0$,其中$\boldsymbol{w} = (w_1, \cdots, w_m)^{\mathrm{T}}$,$\boldsymbol{v} = (v_1, \cdots, v_l)^{\mathrm{T}}$称为 K-T **乘子**,$\boldsymbol{w}$和$\boldsymbol{v}$也称为 **Lagrange 乘子**(Lagrange multipliers). 一阶必要条件可以表达为

$$\left.\begin{aligned}
&\nabla_x L(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}) = 0, \\
&g_i(x) \geqslant 0 \quad i = 1, \cdots, m, \\
&h_j(x) = 0 \quad i = 1, \cdots, l, \\
&w_i g_i(x) = 0 \quad i = 1, \cdots, m, \\
&w_i \geqslant 0 \quad i = 1, \cdots, m.
\end{aligned}\right\} \quad (2.10)$$

对于凸规划,下面给出最优解的充分条件.

**定理 2.1.44** 设在式(2.6)中,$f$是凸函数,$g_i (i = 1, \cdots, m)$是凹函数,$h_j (j = 1, \cdots, l)$是线性函数,可行域为$S$,$\bar{x} \in S$,$I = \{i \mid g_i(\bar{x}) = 0\}$,且在$\bar{x}$处 Kuhn-Tucker 必要条件成立,即存在$w_i \geqslant 0 (i \in I)$及$v_j, j = 1, \cdots, l$使得

$$\nabla f(\bar{x}) - \sum_{i \in I} w_i \nabla g_i(\bar{x}) - \sum_{j=1}^{l} v_j \nabla h_j(\bar{x}) = 0,$$

则$\bar{x}$是整体最优解.

**例 2.1.45** 求下列问题的最优解:

$$\begin{aligned}
\min \quad & (x_1 - 2)^2 + (x_2 - 1)^2; \\
\text{s.t.} \quad & -x_1^2 + x_2 \geqslant 0, \\
& -x_1 - x_2 + 2 \geqslant 0.
\end{aligned}$$

**解** 先求满足 Kuhn-Tucker 必要条件的点，简称 Kuhn-Tucker 点. 目标函数和约束函数的梯度分别是

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 2) \\ 2(x_2 - 1) \end{bmatrix},$$

$$\nabla g_1(\boldsymbol{x}) = \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix},$$

$$\nabla g_2(\boldsymbol{x}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

对于这个问题一阶必要条件(2.10)如下：

$$\begin{aligned}
&2(x_1 - 2) + 2w_1x_1 + w_2 = 0, \\
&2(x_2 - 1) - w_1 + w_2 = 0, \\
&w_1(-x_1^2 + x_2) = 0, \\
&w_2(-x_1 - x_2 + 2) = 0, \\
&-x_1^2 + x_2 \geqslant 0, \\
&-x_1 - x_2 + 2 \geqslant 0, \\
&w_1, w_2 \geqslant 0.
\end{aligned}$$

求解上述问题，得到

$$x_1 = 1,\ x_2 = 1,\ w_1 = \frac{2}{3},\ w_2 = \frac{2}{3}.$$

$\bar{\boldsymbol{x}} = (1,1)^{\mathrm{T}}$ 是 Kuhn-Tucker 点。由于本例是凸规划，根据 2.144，$\bar{\boldsymbol{x}}$ 是这个问题的整体最优解. 参看图 2.1.

最后，给出局部最优解的二阶必要条件和充分条件.

**定理 2.1.46** 设 $\bar{\boldsymbol{x}}$ 是问题(2.6)的局部最优解，$f, g_i\ (i = 1,\cdots,m)$ 和 $h_j\ (j=1,\cdots,l)$ 二阶连续可微，并存在满足(2.10)的乘子 $\boldsymbol{w} = (w_1,\cdots,w_m)$ 和 $\boldsymbol{v} = (v_1,\cdots,v_l)$，再设在点 $\bar{\boldsymbol{x}}$ **约束规格**(constraint qualification)$\overline{G} = \overline{T}$成立，则对每一个向量 $\boldsymbol{d} \in \overline{G}$，都有

$$\boldsymbol{d}^{\mathrm{T}} \nabla_x^2 L(\bar{\boldsymbol{x}}, \boldsymbol{w}, \boldsymbol{v}) \boldsymbol{d} \geqslant 0.$$

其中

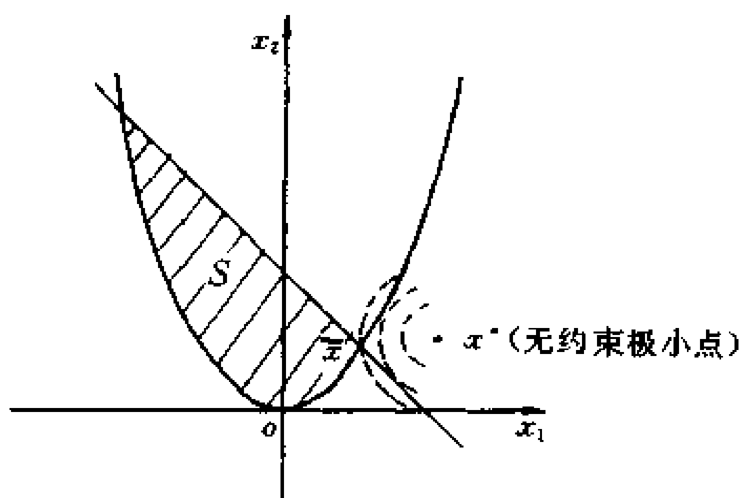



图 2.1

$$\nabla_x^2 L(\bar{x},w,v) = \nabla^2 f(\bar{x}) - \sum_{i=1}^{m} w_i \nabla^2 g_i(\bar{x}) - \sum_{j=1}^{l} v_j \nabla^2 h_j(\bar{x})$$

是 Lagrange 函数 $L(x,w,v)$ 在点 $\bar{x}$ 关于 $x$ 的 Hesse 矩阵.

为解释定理 2.1.46 中的 $\overline{G}$ 和 $\overline{T}$，运用在点 $\bar{x}$ 处起作用约束定义一个集合

$$\overline{S} = \left\{ x \left| \begin{array}{l} g_i(x) = 0, i \in I \text{ 且 } w_i > 0 \\ g_i(x) \geqslant 0, i \in I \text{ 且 } w_i = 0 \\ h_j(x) = 0, j = 1,\cdots,l \end{array} \right. \right\}.$$

定理中的集合 $\overline{G}$ 是集合 $\overline{S}$ 在 $\bar{x}$ 的线性化锥，其定义为

$$\overline{G} = \left\{ d \left| \begin{array}{l} \nabla g_i(\bar{x})^{\mathrm{T}} d = 0, i \in I \text{ 且 } w_i > 0 \\ \nabla g_i(\bar{x}) d \geqslant 0, i \in I \text{ 且 } w_i = 0 \\ \nabla h_j(\bar{x})^{\mathrm{T}} d = 0, j = 1,\cdots,l \end{array} \right. \right\}. \qquad (2.11)$$

$\overline{T}$ 是集合 $\overline{S}$ 在 $\bar{x}$ 的切锥. 定理 2.1.46 中所用约束规格即：$\overline{S}$ 的线性化锥与切锥相等. 这个约束规格不便于检验，可代之以其他约束规格，比如，假设向量组 $\nabla g_i(\bar{x})$ $(i\in I)$，$\nabla h_j(\bar{x})$ $(j=1,\cdots,l)$ 线性无关. 可以证明，若后者成立，则定理 2.1.46 中所用约束规格也

成立.

**定理 2.1.47** 设在问题(2.6)中,$f, g_i (i=1,\cdots,m)$和$h_j (j=1,\cdots,l)$二次连续可微,$\bar{x}$为可行点,存在乘子$\bar{w}=(\bar{w}_1,\cdots,\bar{w}_m)$和$\bar{v}=(\bar{v}_1,\cdots,\bar{v}_l)$使条件(2.10)成立,且对每个向量$d\in G$,都有

$$d^T\nabla_x^2 L(\bar{x},\bar{w},\bar{v})d > 0,$$

则$\bar{x}$是严格的局部最优解.

其中集合$G$是由$\bar{G}$中非零向量组成的集合.$\bar{G}$的定义如(2.11).

**例 2.1.48** 考虑下列非线性规划

$$\begin{aligned} \min \quad & x_1^2 + (x_2 - 2)^2; \\ \text{s.t.} \quad & \beta x_1^2 - x_2 = 0. \end{aligned}$$

其中$\beta$为某个实数.讨论点$x^{(0)}=(0,0)^T$是否为局部最优解.

**解** 目标函数$f(x)$和约束函数$h(x)$在$x^{(0)}$点的梯度分别为

$$\nabla f(x^{(0)}) = \begin{bmatrix} 0 \\ -4 \end{bmatrix}, \nabla h(x^{(0)}) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

设

$$\begin{bmatrix} 0 \\ -4 \end{bmatrix} - v\begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

得到$v=4$,可知$x^{(0)}$是 Kuhn-Tucker 点. Lagrange 函数为

$$L(x,v) = x_1^2 + (x_2 - 2)^2 - v(\beta x_1^2 - x_2),$$

它关于$x$的 Hesse 矩阵是

$$\nabla_x^2 L = \begin{bmatrix} 2 - 2\beta v & 0 \\ 0 & 2 \end{bmatrix}.$$

在点$x^{(0)}$处,有

$$\nabla_x^2 L(x^{(0)},v) = \begin{bmatrix} 2 - 8\beta & 0 \\ 0 & 2 \end{bmatrix}.$$

集合$\bar{G}$中的元素$d=(d_1,0)^T$,其中$d_1$可取任何实数.这时有

$$d^T \nabla_x^2 L(x^{(0)}, v) d = 2(1-4\beta) d_1^2.$$

当 $\beta < 1/4$ 时，对每一个向量 $d \in G$，有 $d^T \nabla_x^2 L(x^{(0)}, v) d > 0$，因此 $x^{(0)} = (0,0)^T$ 是局部最优解.

当 $\beta > 1/4$ 时，对每一个向量 $d \in G$，有 $d^T \nabla_x^2 L(x^{(0)}, v) d < 0$，此时在点 $x^{(0)}$ 不满足局部最优解的二阶必要条件，因此 $x^{(0)} = (0,0)^T$ 不是局部最优解.

当 $\beta = 1/4$ 时，利用二阶条件得不出结论. 可用其他方法进行判断. 这时原问题为：

$$\begin{aligned} \min \quad & x_1^2 + (x_2 - 2)^2; \\ \text{s.t.} \quad & \frac{1}{4} x_1^2 - x_2 = 0. \end{aligned}$$

利用约束条件，从目标函数中消去一个变量，把约束问题转化成无约束问题：

$$\min \quad 4x_2 + (x_2 - 2)^2.$$

易知 $x^{(0)} = (0,0)^T$ 是局部最优解.

可见 $x^{(0)}$ 是否为局部最优解，与参数 $\beta$ 的取值有关. 当 $\beta \leqslant 1/4$ 时，是局部最优解；当 $\beta > 1/4$ 时，不是局部最优解.

### 2.1.5 一维搜索

求解非线性规划所用的计算方法，最常见的是迭代下降算法. 其一般步骤是，得到点 $x^{(k)}$ 后，按某种规则确定一个方向 $d^{(k)}$，从 $x^{(k)}$ 出发沿此方向在直线(或射线)上求目标函数的极小点，从而得到 $x^{(k)}$ 的后继点 $x^{(k+1)}$. 再从 $x^{(k+1)}$ 出发重复以上步骤，直至求得问题的解. 这种方法称为一维搜索，或**线搜索**(line search).

一维搜索可归结为单变量函数的极小化问题. 设目标函数为 $f(x)$，过点 $x^{(k)}$ 沿方向 $d^{(k)}$ 的直线可用点集来表示，记作

$$L = \{x \mid x = x^{(k)} + \lambda d^{(k)}, -\infty < \lambda < \infty\}.$$

求 $f(x)$ 在直线 $L$ 上的极小点转化为求一元函数

$$\varphi(\lambda) = f(x^{(k)} + \lambda d^{(k)})$$

的极小点.

以下考虑的问题一律记作

$$\min \quad f(x) \quad x \in E^1. \tag{2.12}$$

## 1　黄金分割法(0.618 法)

先介绍**黄金分割法**(golden section method)原理.

设 $f$ 是定义在区间$[a,b]$上的单变量 $x$ 的函数. 假定 $f$ 是单峰的,即有唯一的极小点. 在此假设下,可以选择两个试探点,使包含极小点的区间缩短. 比如取点 $\lambda_1, \mu_1 \in (a,b)$,令 $\lambda_1 < \mu_1$,极小点记作 $\bar{x}$. 必有下列两种情形之一:

(1) 如果 $f(\lambda_1) > f(\mu_1)$,则 $\bar{x} \in [\lambda_1, b]$;

(2) 如果 $f(\lambda_1) \leqslant f(\mu_1)$,则 $\bar{x} \in [a, \mu_1]$.

黄金分割法的基本思想是,通过选择试探点缩短包含极小点的区间,不定区间缩短到一定程度时,区间内任一点都可作为极小点的近似.

初始区间记作$[a_1, b_1]$,第 $k$ 次迭代时不定区间记作$[a_k, b_k]$. 黄金分割法计算试探点的公式如下:

$$\lambda_k = a_k + 0.382(b_k - a_k), \tag{2.13}$$

$$\mu_k = a_k + 0.618(b_k - a_k). \tag{2.14}$$

运用黄金分割法,第 1 次迭代取两个试探点 $\lambda_1$ 和 $\mu_1$,以后每次迭代中,只需按照公式(2.13)或(2.14)新算一点.

计算步骤:

(1) 置初始区间$[a_1, b_1]$及精度要求 $L > 0$,计算试探点 $\lambda_1$ 和 $\mu_1$,计算函数值 $f(\lambda_1)$和 $f(\mu_1)$. 计算公式是:

$$\lambda_1 = a_1 + 0.382(b_1 - a_1),$$

$$\mu_1 = a_1 + 0.618(b_1 - a_1).$$

令 $k=1$.

(2) 若 $b_k - a_k < L$，则停止计算. 否则，当 $f(\lambda_k) > f(\mu_k)$ 时，转(3)；当 $f(\lambda_k) \leqslant f(\mu_k)$ 时，转(4).

(3) 置 $a_{k+1} = \lambda_k, b_{k+1} = b_k, \lambda_{k+1} = \mu_k, \mu_{k+1} = a_{k+1} + 0.618(b_{k+1} - a_{k+1})$. 计算函数值 $f(\mu_{k+1})$，转(5).

(4) 置 $a_{k+1} = a_k, b_{k+1} = \mu_k, \mu_{k+1} = \lambda_k, \lambda_{k+1} = a_{k+1} + 0.382(b_{k+1} - a_{k+1})$，计算函数值 $f(\lambda_{k+1})$，转(5).

(5) 置 $k := k+1$，返回(2).

**2 Fibonacci 法**

这种方法与黄金分割法类似，也是用于单峰函数，在计算过程中，第 1 次迭代需要计算两个试探点，以后每次迭代只需新算一点，另一试探点取自上次迭代. Fibonacci 法与黄金分割法的主要区别在于区间长度缩短比率不是常数，而是由 Fibonacci 数确定.

**定义 2.1.49** 设有数列 $\{F_k\}$，满足条件：

(1) $F_0 = F_1 = 1$，

(2) $F_{k+1} = F_k + F_{k-1}, k = 1, 2, \cdots$，

则称 $\{F_k\}$ 为 Fibonacci 数列.

Fibonacci 数列表如下：

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_k$ | 1 | 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | 55 | 89 | … |

Fibonacci 法在迭代中计算试探点的公式为

$$\lambda_k = a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k), k = 1, \cdots, n-1. \tag{2.15}$$

$$\mu_k = a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k), k = 1, \cdots, n-1. \tag{2.16}$$

其中 $n$ 是计算函数值的次数(不包括初始区间端点的计算)，需要事先给定.

设初始区间长度 $b_1-a_1$ 及精度要求(最终区间长度)$L$ 已知,可以求出计算函数值的次数 $n$(不包括初始区间端点函数值的计算). 令 $b_n-a_n\leqslant L$,即

$$\frac{1}{F_n}(b_1-a_1)\leqslant L,$$

由此推出

$$F_n\geqslant\frac{b_1-a_1}{L}. \tag{2.17}$$

先由(2.17)求出 Fibonacci 数 $F_n$,再根据 $F_n$ 确定计算函数值的次数 $n$.

运用 Fibonacci 法时,应注意下列问题:

由于第 1 次迭代计算两个试探点,以后每次计算一个,这样经过 $n-1$ 次迭代就算完 $n$ 个试探点. 但是,在第 $n-1$ 次迭代中并没有选择新的试探点. 根据公式(2.15)和(2.16),必有

$$\lambda_{n-1}=\mu_{n-1}=\frac{1}{2}(b_{n-1}+a_{n-1}).$$

而 $\lambda_{n-1}$和 $\mu_{n-1}$中的一个取自第 $n-2$ 次迭代中的试探点. 为了在第 $n-1$ 次迭代中能够缩短不定区间,可在第 $n-2$ 次迭代之后,这时已确定出 $\lambda_{n-1}=\mu_{n-1}$,在 $\lambda_{n-1}$的右边或左边取一点,令

$$\lambda_n=\lambda_{n-1},\mu_n=\lambda_{n-1}+\delta,$$

其中辨别常数 $\delta>0$.

计算步骤:

(1) 给定初始区间$[a_1,b_1]$和最终区间长度 $L$. 求计算函数值的次数 $n$,使

$$F_n\geqslant(b_1-a_1)/L.$$

置辨别常数 $\delta>0$. 计算试探点 $\lambda_1$ 和 $\mu_1$,根据公式(2.15)和(2.16)有:

$$\lambda_1=a_1+\frac{F_{n-2}}{F_n}(b_1-a_1),$$

$$\mu_1 = a_1 + \frac{F_{n-1}}{F_n}(b_1 - a_1),$$

计算函数值 $f(\lambda_1)$和 $f(\mu_1)$. 置 $k=1$.

(2) 若 $f(\lambda_k) > f(\mu_k)$,则转(3);否则,即 $f(\lambda_k) \leqslant f(\mu_k)$,则转(4).

(3) 令 $a_{k+1}=\lambda_k, b_{k+1}=b_k, \lambda_{k+1}=\mu_k$,计算试探点 $\mu_{k+1}$:

$$\mu_{k+1} = a_{k+1} + \frac{F_{n-k-1}}{F_{n-k}}(b_{k+1} - a_{k+1}),$$

若 $k=n-2$,则转(6);否则,计算函数值 $f(\mu_{k+1})$,转(5).

(4) 令 $a_{k+1}=a_k, b_{k+1}=\mu_k, \mu_{k+1}=\lambda_k$,计算 $\lambda_{k+1}$:

$$\lambda_{k+1} = a_{k+1} + \frac{F_{n-k-2}}{F_{n-k}}(b_{k+1} - a_{k+1}),$$

若 $k=n-2$,则转(6);否则,计算函数值 $f(\lambda_{k+1})$,转(5).

(5) 置 $k := k+1$,转(2).

(6) 令 $\lambda_n=\lambda_{n-1}, \mu_n=\mu_{n-1}+\delta$,计算函数值 $f(\lambda_n)$和 $f(\mu_n)$.

若 $f(\lambda_n) > f(\mu_n)$,则令

$$a_n = \lambda_n, b_n = b_{n-1};$$

若 $f(\lambda_n) \leqslant f(\mu_n)$,则令

$$a_n = a_{n-1}, b_n = \lambda_n.$$

停止计算,极小点含于$[a_n, b_n]$.

### 3 Newton 法

设问题(2.12)中,$f(x)$二次可微.

Newton 法的基本思想是,在估计点 $x_k$ 附近使 $f'(x)$线性化,并求出这个线性函数的零点. 这样就得到一个估计点 $x_{k+1}$. 由此推出迭代公式

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}. \tag{2.18}$$

运用公式(2.18)进行迭代,直至

$$|x_{k+1} - x_k| < \varepsilon \text{ 或 } |f'(x_k)| < \varepsilon$$

时，迭代终止. 其中 $\varepsilon$ 是事先给定的精度.

运用 Newton 法时，初始点的选择十分重要，若初始点靠近极小点，则很快收敛；如果初始点远离极小点，迭代产生的点列可能不收敛于极小点.

**定理 2.1.50** 设 $f(x)$ 存在连续三阶导数，$\bar{x}$ 满足 $f'(\bar{x})=0$，$f''(\bar{x})\neq 0$，初始点 $x_1$ 充分接近 $\bar{x}$，则 Newton 法产生的序列 $\{x_k\}$ 至少以 2 阶收敛速率收敛于 $\bar{x}$.

定理 2.1.50 用到**收敛阶**的概念，定义如下：

**定义 2.1.51** 设存在一序列 $\{\boldsymbol{x}^{(k)}\}$，$\boldsymbol{x}^{(k)}\in E^n$，收敛于点 $\bar{\boldsymbol{x}}$，并且对所有充分大的 $k$ 有 $\boldsymbol{x}^{(k)}\neq\bar{\boldsymbol{x}}$. 如果存在数 $p$ 和 $\alpha\neq 0$，使得

$$\lim_{k\to\infty}\frac{\|\boldsymbol{x}^{(k+1)}-\bar{\boldsymbol{x}}\|}{\|\boldsymbol{x}^{(k)}-\bar{\boldsymbol{x}}\|^p}=\alpha,$$

则称序列 $\{\boldsymbol{x}^{(k)}\}$ 为 $p$ 阶收敛.

如果 $p=1$，则称 $\{\boldsymbol{x}^{(k)}\}$ 的收敛速率是线性的.

**4 割线法(secant method)**

这种方法的基本思想是，用割线逼近目标函数的导函数曲线 $y=f'(x)$，把割线的零点作为目标函数驻点的估计，如图 2.2 所示.

由此推出迭代公式

$$x_{k+1}=x_k-\frac{x_k-x_{k-1}}{f'(x_k)-f'(x_{k-1})}f'(x_k). \tag{2.19}$$

用这个公式进行迭代，得到序列 $\{x_k\}$，可以证明，在一定条件下，这个序列收敛于解.

**定理 2.1.52** 设 $f(x)$ 存在连续三阶导数，$\bar{x}$ 满足 $f'(\bar{x})=0$，$f''(\bar{x})\neq 0$，若 $x_1$ 和 $x_2$ 充分接近 $\bar{x}$，则割线法产生的序列 $\{x_k\}$ 收敛于 $\bar{x}$，且收敛阶为 1.618.

**5 抛物线法(parabolic method)**

抛物线法的基本思想是，在极小点附近，用二次三项式逼近目

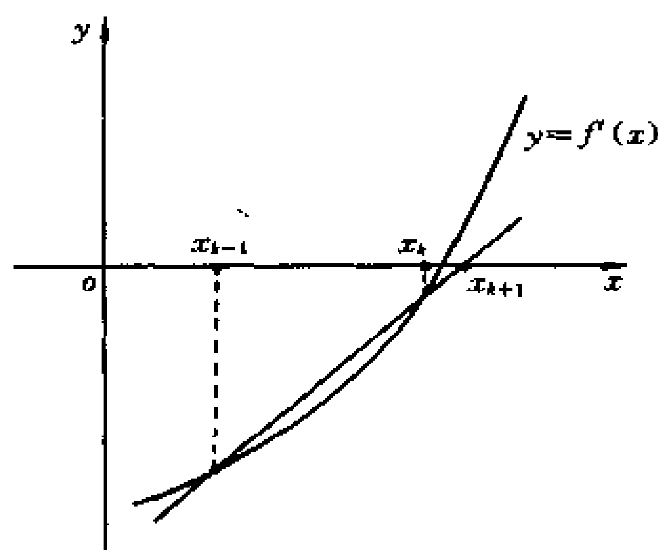



图 2.2

标函数 $f(x)$，令此二次三项式与 $f(x)$ 在三点 $x_1<x_2<x_3$ 处有相同的函数值，并假设 $f(x_1)>f(x_2)$，$f(x_2)<f(x_3)$，求二次三项式的极小点，作为 $f(x)$ 极小点的一个估计，从而得到迭代公式如下：

$$\bar{x}_k = \frac{(x_2^2 - x_3^2)f(x_1) + (x_3^2 - x_1^2)f(x_2) + (x_1^2 - x_2^2)f(x_3)}{2[(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)]}. \tag{2.20}$$

$\bar{x}_k$ 作为 $f(x)$ 的极小点的一个估计. 再从 $x_1,x_2,x_3,\bar{x}_k$ 中选择目标函数值最小的点及其左、右两点，给予相应的下标，即中间一点记作 $x_2$，左、右两点分别记作 $x_1$ 和 $x_3$，代入公式(2.20)，求出极小点的新的估计值 $\bar{x}_{k+1}$，依此类推，产生点列 $\{\bar{x}_k\}$. 在一定的条件下，这个点列收敛于问题的解，其收敛阶为 1.3. 在实际应用中，只要满足精度要求即可. 一般用目标函数的下降量或位移来控制，即当

$$|f(\bar{x}_{k+1}) - f(\bar{x}_k)| < \epsilon,$$

或者当

$$\| \bar{x}_{k+1} - \bar{x}_k \| < \delta$$

时，终止迭代. 其中 $\epsilon,\delta$ 为事先给定的允许误差.

**注意**，三个初始点 $x_1<x_2<x_3$ 的选择，必须满足

$$f(x_1) > f(x_2), f(x_2) < f(x_3),$$

这样才能保证极小点在区间 $(x_1,x_3)$ 内.

**6 三次插值法(cubic interpolation method)**

二点三次插值法的基本思想是，取两个初点 $x_1$ 和 $x_2(x_1<x_2)$，使得 $f'(x_1)<0$ 及 $f'(x_2)>0$，构造一个三次多项式，使它与 $f(x)$ 在 $x_1$ 和 $x_2$ 处有相同的函数值及导数值，用这个多项式的极小点作为 $f(x)$ 极小点的估计值. 用此方法的迭代公式为

$$\bar{x} = x_1 + (x_2 - x_1)\left(1 - \frac{f'(x_2) + v + w}{f'(x_2) - f'(x_1) + 2w}\right), \tag{2.21}$$

其中 $v$ 和 $w$ 利用下列公式计算：

$$s = \frac{3[f(x_2) - f(x_1)]}{x_2 - x_1}. \tag{2.22}$$

$$v = s - f'(x_1) - f'(x_2). \tag{2.23}$$

$$w^2 = v^2 - f'(x_1)f'(x_2). \tag{2.24}$$

若 $|f'(\bar{x})|$ 充分小，则 $\bar{x}$ 可作为 $f(x)$ 的可接受的极小点；否则，从 $x_1,x_2$ 和 $\bar{x}$ 中确定两个插值点，再利用上述公式进行计算.

计算步骤：

(1) 给定初始点 $x_1,x_2$，计算 $f(x_1),f(x_2),f'(x_1),f'(x_2)$，要求满足条件：$x_2>x_1,f'(x_1)<0,f'(x_2)>0$，给定允许误差 $\delta>0$.

(2) 按照公式(2.22)～(2.24)和(2.21)计算 $s,v,w$ 和 $\bar{x}$.

(3) 若 $|x_2-x_1|\leqslant\delta$，则停止计算，得到点 $\bar{x}$；否则，进行(4).

(4) 计算 $f(\bar{x}),f'(\bar{x})$. 若 $f'(\bar{x})=0$，则停止计算，得到点 $\bar{x}$.

若 $f'(\bar{x})<0$，则令 $x_1=\bar{x},f(x_1)=f(\bar{x}),f'(x_1)=f'(\bar{x})$，转(2).

若 $f'(\bar{x})>0$，则令 $x_2=\bar{x},f(x_2)=f(\bar{x}),f'(x_2)=f'(\bar{x})$，

转(2).

## 2.2 无约束非线性规划

考虑问题：

$$\min \quad f(\boldsymbol{x}), \boldsymbol{x} \in E^n \tag{2.25}$$

求 $f(\boldsymbol{x})$在 $E^n$ 中的极小点，一般通过一系列一维搜索来实现，其核心问题是选择搜索方向. 搜索方向不同则形成不同的最优化方法.

无约束问题的最优化方法一般分作两类：一类在计算过程中使用导数，可称为**使用导数的最优化方法**；另一类在计算过程中只用到目标函数，通常称为**直接方法**. 本节中，首先给出使用导数的方法，然后介绍直接方法.

### 2.2.1 最速下降法

**最速下降法**(steepest descent method)由法国数学家 Cauchy 于 1847 年首先提出. 这种方法，在每次迭代中，沿最速下降方向(负梯度方向)进行搜索. 迭代公式为

$$\left.\begin{aligned} &\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}, \\ &\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}), \\ &\lambda_k: f(\boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}) = \min_{\lambda \geq 0} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)}). \end{aligned}\right\} \tag{2.26}$$

算法如下：

(1) 给定初点 $\boldsymbol{x}^{(1)} \in E^n$，允许误差 $\varepsilon > 0$，置 $k=1$.

(2) 求搜索方向 $\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})$.

(3) 若 $\|\boldsymbol{d}^{(k)}\| < \varepsilon$，则停止计算；否则，从 $\boldsymbol{x}^{(k)}$出发，沿 $\boldsymbol{d}^{(k)}$进行一维搜索，求 $\lambda_k$，使得

$$f(\boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}) = \min_{\lambda \geq 0} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)}).$$

(4) 令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$，置 $k:=k+1$，转(2).

**例 2.2.1** 用最速下降法解下列问题

$$\min \quad f(\boldsymbol{x}) = 2x_1^2 + x_2^2,$$

初点 $\boldsymbol{x}^{(1)}=(1,1)^{\mathrm{T}}$，$\epsilon=\frac{1}{10}$.

**解** 第 1 次迭代

目标函数 $f(\boldsymbol{x})$ 在点 $\boldsymbol{x}$ 处的梯度及搜索方向为

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 4x_1 \\ 2x_2 \end{bmatrix}, \boldsymbol{d}^{(1)} = -\nabla f(\boldsymbol{x}^{(1)}) = \begin{bmatrix} -4 \\ -2 \end{bmatrix},$$

$\|\boldsymbol{d}\|=2\sqrt{5}>\frac{1}{10}$. 从 $\boldsymbol{x}^{(1)}=(1,1)^{\mathrm{T}}$ 出发，沿方向 $\boldsymbol{d}^{(1)}$ 进行一维搜索，求得步长 $\lambda_1=5/18$. 在直线上的极小点

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \lambda_1\boldsymbol{d}^{(1)} = \begin{bmatrix} -\frac{1}{9} \\ \frac{4}{9} \end{bmatrix}.$$

第 2 次迭代

$f(\boldsymbol{x})$ 在点 $\boldsymbol{x}^{(2)}$ 处的最速下降方向为

$$\boldsymbol{d}^{(2)} = -\nabla f(\boldsymbol{x}^{(2)}) = \begin{bmatrix} \frac{4}{9} \\ -\frac{8}{9} \end{bmatrix},$$

$\|\boldsymbol{d}^{(2)}\|=\frac{4}{9}\sqrt{5}>\frac{1}{10}$. 不满足精度要求. 从 $\boldsymbol{x}^{(2)}$ 出发，沿方向 $\boldsymbol{d}^{(2)}$ 进行一维搜索，得到步长 $\lambda_2=5/12$，沿此方向得到的极小点

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \lambda_2\boldsymbol{d}^{(2)} = \frac{2}{27}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

第 3 次迭代

$f(\boldsymbol{x})$ 在点 $\boldsymbol{x}^{(3)}$ 处的最速下降方向

$$d^{(3)} = -\nabla f(x^{(3)}) = \frac{4}{27}\begin{bmatrix} -2 \\ -1 \end{bmatrix},$$

由于 $\| d^{(3)} \| > \frac{1}{10}$，不满足精度要求. 再从 $x^{(3)}$ 出发，沿 $d^{(3)}$ 作一维搜索，得到 $\lambda_3 = 5/18$.

$$x^{(4)} = x^{(3)} + \lambda_3 d^{(3)} = \frac{2}{243}\begin{bmatrix} -1 \\ 4 \end{bmatrix}.$$

这时有 $\| \nabla f(x^{(4)}) \| < \frac{1}{10}$，已满足精度要求，得到问题的近似解

$$\bar{x} = \frac{2}{243}\begin{bmatrix} -1 \\ 4 \end{bmatrix}.$$

实际上，问题的最优解 $x^* = (0,0)^T$.

最速下降法在一定条件下是收敛的.

**定理 2.2.2** 设 $f(x)$ 是连续可微实函数，解集合 $\Omega = \{\bar{x} \mid \nabla f(\bar{x}) = 0\}$，最速下降算法产生的序列 $\{x^{(k)}\}$ 含于某个紧集，则序列 $\{x^{(k)}\}$ 的每个聚点 $\hat{x} \in \Omega$.

最速下降法产生的序列是**线性收敛**(linear convergence)的，而且收敛性质与极小点处 Hesse 矩阵 $\nabla^2 f(\bar{x})$ 的特征值有关.

**定理 2.2.3** 设 $f(x)$ 存在连续二阶偏导数，$\bar{x}$ 是局部极小点，Hesse 矩阵 $\nabla^2 f(\bar{x})$ 的最小特征值 $a > 0$，最大特征值为 $A$，算法产生的序列 $\{x^{(k)}\}$ 收敛于 $\bar{x}$，则目标函数值的序列 $\{f(x^{(k)})\}$ 以不大于 $\left(\frac{A-a}{A+a}\right)^2$ 的收敛比线性地收敛于 $f(\bar{x})$.

上述定理中，若令 $r = A/a$，则

$$\left(\frac{A-a}{A+a}\right)^2 = \left(\frac{r-1}{r+1}\right)^2 < 1,$$

$r$ 是对称正定矩阵 $\nabla^2 f(\bar{x})$ 的**条件数**(condition number). 这个定理表明，条件数越小，收敛越快；条件数越大，收敛越慢.

用最速下降法极小化目标函数时，相邻两个搜索方向是正交

的，因此迭代产生的序列$\{x^{(k)}\}$所循路径是“之”字形的. 当 $x^{(k)}$ 接近极小点 $\bar{x}$ 时，每次迭代移动的步长很小，这样就呈现出锯齿现象. 当 Hesse 矩阵$\nabla^2 f(\bar{x})$的条件数比较大时，锯齿现象的影响尤为严重. 对此可作如下分析：

粗略地讲，在极小点附近，目标函数一般可用二次函数近似，其等值面接近椭球面，长轴和短轴分别位于对应最小特征值和最大特征值的特征向量方向，其大小与特征值的平方根成反比. 最小特征值与最大特征值相差越大，椭球面越扁，这就使得一维搜索沿着“斜长谷”进行. 因此，当条件数很大时，要使迭代点充分接近极小点，就需要走很大的弯路，计算效率很低.

最速下降方向反映了目标函数的一种局部性质. 从局部看，最速下降方向确是目标函数值下降最快的方向，选择这样的方向进行搜索是有利的. 但从全局看，由于锯齿现象的影响，即使向着极小点移近不太大的距离，也要经历不小的“弯路”，因此收敛速率大为减慢. 最速下降法并不是收敛最快的方法，相反，从全局看，它的收敛是比较慢的. 因此，最速下降法一般适用于计算过程的前期迭代，或作为间插步骤. 当接近极小点时，再使用此法，试图达到迭代的终止，并非有利.

### 2.2.2 Newton 法

#### 1 Newton 法原理

设问题(2.25)中，$f(x)$为二次可微实函数. Newton 法的基本思想是，用一个二次函数局部地近似 $f(x)$，然后求出此近似函数的极小点，从而得到迭代公式：

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}), \qquad (2.27)$$

其中$\nabla^2 f(x^{(k)})^{-1}$是 Hesse 矩阵$\nabla^2 f(x^{(k)})$的逆矩阵. 这样，知道 $x^{(k)}$后，算出在这一点处目标函数的梯度和 Hesse 逆矩阵，代入(2.27)，便得到后继点 $x^{(k+1)}$，用 $k+1$ 代替 $k$，再用(2.27)计算，又

得到 $x^{(k+1)}$的后继点. 依此类推，产生点列$\{x^{(k)}\}$. 在适当条件下，这个序列收敛.

**定理 2.2.4** 设 $f(x)$为二次连续可微函数，$x\in E^n$，$\bar{x}$ 满足 $\nabla f(\bar{x})=0$，且$\nabla^2 f(\bar{x})^{-1}$存在. 又设初始点 $x^{(1)}$充分接近 $\bar{x}$，使得存在 $k_1,k_2>0$，满足 $k_1k_2<1$，且对每一个

$$x \in \{x \mid \| x - \bar{x} \| \leqslant \| x^{(1)} - \bar{x} \| \}$$

成立

(1) $\| \nabla^2 f(x)^{-1} \| \leqslant k_1$，

(2) $\dfrac{\| \nabla f(\bar{x}) - \nabla f(x) - \nabla^2 f(x)(\bar{x}-x) \|}{\| \bar{x}-x \|} \leqslant k_2$，

则 Newton 法产生的序列收敛于 $\bar{x}$.

当 Newton 法收敛时，有下列关系：

$$\| x^{(k+1)} - \bar{x} \| \leqslant C \| x^{(k)} - \bar{x} \|^2 \tag{2.28}$$

其中 $C$ 是某个常数. 由此可知，Newton 法至少二阶收敛. 特别地，$f(x)$为二次凸函数时，运用 Newton 法，经 1 次迭代达极小点.

**注意**，当初始点远离极小点时，Newton 法可能不收敛. 原因之一，Newton 方向

$$d = - \nabla^2 f(x)^{-1} \nabla f(x) \tag{2.29}$$

不一定是下降方向，经迭代，目标函数值可能上升. 此外，即使目标函数值下降，得到的点 $x^{(k+1)}$也不一定是沿 Newton 方向的最好点或极小点. 因此有阻尼 Newton 法.

**2 阻尼 Newton 法**(damped Newton's method)

阻尼 Newton 法与原始 Newton 法的区别在于增加了沿 Newton 方向的一维搜索，其迭代公式是：

$$\left.\begin{aligned} & x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}, \\ & d^{(k)} = - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}), \\ & \lambda_k :\ f(x^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda} f(x^{(k)} + \lambda d^{(k)}) \end{aligned}\right\} \tag{2.30}$$

计算步骤：

(1) 给定初始点 $\boldsymbol{x}^{(1)}$，允许误差 $\varepsilon>0$，置 $k=1$.

(2) 计算 $\nabla f(\boldsymbol{x}^{(k)})$，$\nabla^2 f(\boldsymbol{x}^{(k)})^{-1}$.

(3) 若 $\|\nabla f(\boldsymbol{x}^{(k)})\|<\varepsilon$，则停止迭代；否则，令

$$\boldsymbol{d}^{(k)}=-\nabla^2 f(\boldsymbol{x}^{(k)})^{-1}\nabla f(\boldsymbol{x}^{(k)}).$$

(4) 从 $\boldsymbol{x}^{(k)}$ 出发，沿 $\boldsymbol{d}^{(k)}$ 方向作一维搜索：

$$\min_{\lambda} f(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)})=f(\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}),$$

令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$.

(5) 置 $k:=k+1$，转(2).

由于阻尼 Newton 法含有一维搜索，因此，每次迭代目标函数值一般有所下降(决不会上升). 可以证明，阻尼 Newton 法在适当条件下具有全局收敛性，且为二阶收敛.

原始 Newton 法和阻尼 Newton 法有共同缺点：一是可能出现 Hesse 矩阵奇异的情形，因此不能确定后继点；二是即使 $\nabla^2 f(\boldsymbol{x})$ 非奇异，也未必正定，因而 Newton 方向不一定是下降方向，这就可能导致算法失效.

**例 2.2.5** 用阻尼 Newton 法求解下列问题：

$$\min \quad f(\boldsymbol{x}) \triangleq x_1^4+x_1x_2+(1+x_2)^2.$$

取初始点 $\boldsymbol{x}^{(1)}=(0,0)^{\mathrm{T}}$. 在点 $\boldsymbol{x}^{(1)}$ 处，函数 $f(\boldsymbol{x})$ 的梯度和 Hesse 矩阵分别为

$$\nabla f(\boldsymbol{x}^{(1)})=\begin{bmatrix}0\\2\end{bmatrix},\nabla^2 f(\boldsymbol{x}^{(1)})=\begin{bmatrix}0&1\\1&2\end{bmatrix}.$$

牛顿方向

$$\boldsymbol{d}^{(1)}=-\nabla^2 f(\boldsymbol{x}^{(1)})^{-1}\nabla f(\boldsymbol{x}^{(1)})=\begin{bmatrix}-2\\0\end{bmatrix}.$$

从 $\boldsymbol{x}^{(1)}$ 出发沿 $\boldsymbol{d}^{(1)}$ 方向作一维搜索，得到步长 $\lambda_1=0$. 因此经一次迭代没有产生新点，仍然得到 $\boldsymbol{x}^{(1)}=(0,0)^{\mathrm{T}}$. 显然，$\boldsymbol{x}^{(1)}$ 并不是问题的极小点. 可见，从 $\boldsymbol{x}^{(1)}$ 出发，用阻尼 Newton 法求不出极小点. 其原

因在于 Hesse 矩阵$\nabla^2 f(\boldsymbol{x}^{(1)})$非正定.

**3　Marquardt-Levenberg 方法**

为了使 Newton 法适用于$\nabla^2 f(\boldsymbol{x}^{(k)})$不是正定的情形,人们对于 Newton 法提出了若干修正方法. Marquardt-Levenberg 方法就是其中之一. 这种方法的迭代公式是:

$$\left.\begin{aligned} &\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}, \\ &\boldsymbol{d}^{(k)} = -\boldsymbol{G}_k^{-1}\nabla f(\boldsymbol{x}^{(k)}), \\ &\boldsymbol{G}_k = \nabla^2 f(\boldsymbol{x}^{(k)}) + \beta_k \boldsymbol{Q}_k, \\ &\lambda_k:\ f(\boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}) = \min_{\lambda\geqslant 0} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)}), \end{aligned}\right\} \quad (2.31)$$

其中 $\beta_k$ 是一个非负数,$\boldsymbol{Q}_k$ 是给定的正定矩阵.

迭代公式中,当 $\beta_k$ 取得足够小时,$\boldsymbol{G}_k$ 的第二项 $\beta_k\boldsymbol{Q}_k$ 几乎不起作用,因此方法近似于 Newton 法;当 $\beta_k$ 取得足够大时,$\boldsymbol{G}_k$ 中主要是 $\beta_k\boldsymbol{Q}_k$ 起作用,又若 $\boldsymbol{Q}_k$ 取为单位矩阵,此时近似于最速下降法. 当$\nabla^2 f(\boldsymbol{x}^{(k)})$非正定时,只要 $\beta_k$ 取得足够大,便可保证 $\boldsymbol{G}_k$ 正定,从而 $\boldsymbol{d}^{(k)}$为下降方向.

计算步骤:

(1) 给定 $\boldsymbol{x}^{(1)}, \beta_1, \boldsymbol{Q}, \epsilon, \alpha>1$,置 $k=1$.

(2) 求出$\nabla f(\boldsymbol{x}^{(k)}), \nabla^2 f(\boldsymbol{x}^{(k)})$,置 $\beta=\beta_1$.

(3) 令 $\boldsymbol{G}=\nabla^2 f(\boldsymbol{x}^{(k)})+\beta\boldsymbol{Q}$.

(4) 若 $\boldsymbol{G}^{-1}$存在,则求出 $\boldsymbol{G}^{-1}$,转向(5);若 $\boldsymbol{G}^{-1}$不存在,置 $\beta := \alpha\beta$,转(3).

(5) 求 $\boldsymbol{d}^{(k)}=-\boldsymbol{G}^{-1}\nabla f(\boldsymbol{x}^{(k)})$.

(6) 求 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$,其中 $\lambda_k$ 满足:

$$f(\boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}) = \min_{\lambda} f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)}).$$

若 $\lambda_k=0$,则置 $\beta := \alpha\beta$,转(3);如果 $\lambda_k\neq 0$,则转(7).

(7) 若 $\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| \leqslant \epsilon$,则求得最优解 $\boldsymbol{x}^* = \boldsymbol{x}^{(k+1)}$;若 $\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| > \epsilon$,则置 $k := k+1$,转(2).

### 2.2.3 共轭梯度法

#### 1 共轭方向

**共轭梯度法**(conjugate gradient method)是以**共轭方向**(conjugate direction)作为搜索方向的一类算法,先给出有关概念及性质.

**定义 2.2.6** 设 $A$ 是 $n\times n$ 对称正定矩阵,若 $E^n$ 中的两个方向 $d^{(1)}$和 $d^{(2)}$满足

$$d^{(1)\mathrm{T}}Ad^{(2)}=0,$$

则称这两个方向关于 $A$ 共轭,或称它们关于 $A$ 正交.

若 $d^{(1)},d^{(2)},\cdots,d^{(k)}$是 $E^n$ 中 $k$ 个方向,它们两两关于 $A$ 共轭,即满足

$$d^{(i)\mathrm{T}}Ad^{(j)}=0,i\neq j,\quad i,j=1,\cdots,k,$$

则称这组方向是 $A$ 共轭的,或称它们为 $A$ 的 $k$ 个共轭方向.

在上述定义中,如果 $A$ 为单位矩阵,则两个方向关于 $A$ 共轭等价于两个方向正交.因此共轭是正交概念的推广.实际上,如果 $A$ 是一般的对称正定矩阵,$d^{(i)}$和 $d^{(j)}$关于 $A$ 共轭,也就是方向 $d^{(i)}$和方向 $Ad^{(j)}$正交.

共轭方向具有下列性质:

**定理 2.2.7** 设 $A$ 是 $n$ 阶对称正定矩阵,$d^{(1)},d^{(2)},\cdots,d^{(k)}$是 $k$ 个 $A$ 共轭的非零向量,则这组向量线性无关.

**定理 2.2.8** 设有二次函数

$$f(x)=\frac{1}{2}x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c$$

其中 $A$ 是 $n$ 阶对称正定矩阵,$d^{(1)},d^{(2)},\cdots,d^{(k)}$是 $A$ 共轭的非零向量.以任意 $x^{(1)}\in E^n$ 为初始点,依次沿方向 $d^{(1)},d^{(2)},\cdots,d^{(k)}$进行一维搜索,得到点 $x^{(2)},x^{(3)},\cdots,x^{(k+1)}$,则 $x^{(k+1)}$是函数 $f(x)$在**线性流形**(linear manifold)$x^{(1)}+\mathscr{B}_k$ 上的唯一极小点.特别,当 $k=n$ 时,

$\boldsymbol{x}^{(n+1)}$是函数 $f(\boldsymbol{x})$在 $E^n$ 上的唯一极小点. 其中

$$\mathscr{B}_k = \{\boldsymbol{x} \mid \boldsymbol{x} = \sum_{j=1}^{k} \lambda_j \boldsymbol{d}^{(j)}, \lambda_j \in (-\infty, +\infty)\} \tag{2.32}$$

是 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(k)}$生成的子空间.

上述定理称为**扩张子空间定理**(expanding subspace theorem). 它表明,对于二次凸函数,若沿一组共轭方向搜索,经有限步迭代必达到极小点. 这种性质称为**二次终止性**(quadratic termination). 利用扩张子空间定理 2.2.8,容易证明共轭方向的另一个性质.

**定理 2.2.9** 设 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c$, $\boldsymbol{A}$ 是 $n$ 阶对称正定矩阵,又设 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(k)}$是一组 $\boldsymbol{A}$ 共轭的非零向量,$\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)} \in E^n$ 为任意两点,从 $\boldsymbol{x}^{(0)}$出发,依次沿此 $k$ 个方向搜索,得到在流形 $\boldsymbol{x}^{(0)} + \mathscr{B}_k$ 上的极小点 $\boldsymbol{x}^{(a)}$,从 $\boldsymbol{x}^{(1)}$出发,依次沿这 $k$ 个方向搜索,得到在流形 $\boldsymbol{x}^{(1)} + \mathscr{B}_k$ 上的极小点 $\boldsymbol{x}^{(b)}$,则

$$\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(k)}, \boldsymbol{d}^{(k+1)} = \boldsymbol{x}^{(b)} - \boldsymbol{x}^{(a)}$$

是 $\boldsymbol{A}$ 共轭的.

定理中的 $\mathscr{B}_k$ 是 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(k)}$生成的子空间.

### 2 共轭梯度法

这是基于共轭方向的一类算法. 为书写简便起见,用 $g_k$ 表示函数 $f(\boldsymbol{x})$在点 $\boldsymbol{x}^{(k)}$的梯度,即令 $g_k \triangleq \nabla f(\boldsymbol{x}^{(k)})$.

共轭梯度法由 Hesteness 和 Stiefel 于 1952 年为求解线性方程组而提出. 后来用于求解无约束最优化问题,它是一种重要的方法.

共轭梯度法的基本思想是把共轭性与最速下降方法相结合,利用已知点处的梯度构造一组共轭方向,并沿此组方向进行搜索,求出目标函数的极小点. 根据共轭方向的基本性质,这种方法具有二次终止性.

先介绍 Fletcher-Reeves (FR)**共轭梯度法**用于二次凸函数的情形. 考虑问题

$$\min \quad f(\boldsymbol{x}) \triangleq \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + c, \tag{2.33}$$

其中 $\boldsymbol{x} \in E^n$, $\boldsymbol{A}$ 是对称正定矩阵, $c$ 是常数. 运用 Fletcher-Reeves 共轭梯度法时, 迭代公式为

$\boldsymbol{x}^{(1)}$给定, 令 $\boldsymbol{d}^{(1)} = -\boldsymbol{g}_1$,

$$\left.\begin{aligned}
&\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}, \\
&\lambda_k = -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}^{(k)}}{\boldsymbol{d}^{(k)\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^{(k)}}, \\
&\boldsymbol{d}^{(k+1)} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}^{(k)}, \\
&\beta_k = \frac{\boldsymbol{d}^{(k)\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_{k+1}}{\boldsymbol{d}^{(k)\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^{(k)}} = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2}.
\end{aligned}\right\} \tag{2.34}$$

计算步骤:

(1) 给定初始点 $\boldsymbol{x}^{(1)}$, 置 $\boldsymbol{d}^{(1)} = -\boldsymbol{g}_1$, $k=1$.

(2) 从 $\boldsymbol{x}^{(k)}$出发, 沿 $\boldsymbol{d}^{(k)}$搜索, 求出 $\lambda_k$. 令

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{d}^{(k)}.$$

(3) 计算 $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}^{(k+1)})$, 若 $\boldsymbol{g}_{k+1} = \boldsymbol{0}$, 则停止计算.

(4) 令

$$\beta_k = \frac{\|\boldsymbol{g}_{k+1}\|^2}{\|\boldsymbol{g}_k\|^2},$$

$$\boldsymbol{d}^{(k+1)} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}^{(k)}.$$

(5) 置 $k := k+1$, 返回(2).

**例 2.2.10** 用 Fletcher-Reeves 共轭梯度法求解下列问题:

$$\min \quad f(\boldsymbol{x}) \triangleq x_1^2 + 2x_2^2,$$

初始点 $\boldsymbol{x}^{(1)} = (5,5)^{\mathrm{T}}$.

**解** 在点 $\boldsymbol{x}$ 处, 目标函数 $f(\boldsymbol{x})$的梯度$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}$.

第 1 次迭代：令

$$d^{(1)} = -g_1 = \begin{bmatrix} -10 \\ -20 \end{bmatrix},$$

从 $x^{(1)}$出发，沿方向 $d^{(1)}$作一维搜索，求得步长

$$\lambda_1 = -\frac{g_1^T d^{(1)}}{d^{(1)T} A d^{(1)}} = \frac{5}{18},$$

$$x^{(2)} = x^{(1)} + \lambda_1 d^{(1)} = \begin{bmatrix} \frac{20}{9} \\ -\frac{5}{9} \end{bmatrix}.$$

第 2 次迭代：在点 $x^{(2)}$处，目标函数的梯度

$$g_2 = \begin{bmatrix} \frac{40}{9} \\ -\frac{20}{9} \end{bmatrix}.$$

构造搜索方向 $d^{(2)}$. 先计算因子 $\beta_1$：

$$\beta_1 = \frac{\| g_2 \|^2}{\| g_1 \|^2} = \frac{4}{81}.$$

再令搜索方向

$$d^{(2)} = -g_2 + \beta_1 d^{(1)} = \frac{100}{81} \begin{bmatrix} -4 \\ 1 \end{bmatrix},$$

从 $x^{(2)}$出发，沿 $d^{(2)}$作一维搜索. 求 $\lambda_2$：

$$\lambda_2 = -\frac{g_2^T d^{(2)}}{d^{(2)T} A d^{(2)}} = \frac{9}{20},$$

$$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

在 $x^{(3)}$处，目标函数的梯度 $g_3 = (0,0)^T$，经二次迭代即达极小点 $x^{(3)} = (0,0)^T$. 取得如此好的结果并非偶然，Fletcher-Reeves 共轭梯度法用于二次凸函数时，有下列结论：

**定理 2.2.11** 对于正定二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}+c$（$\boldsymbol{A}$ 为 $n$ 阶对称正定矩阵），具有精确一维搜索的 Fletcher-Reeves 法在 $m(\leqslant n)$ 次一维搜索后即终止，并且对所有 $i(1\leqslant i\leqslant m)$，有下列关系：

(1) $\boldsymbol{d}^{(i)\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^{(j)}=0,\ j=1,2,\cdots,i-1$；

(2) $\boldsymbol{g}_i^{\mathrm{T}}\boldsymbol{g}_j=0,\ j=1,2,\cdots,i-1$；

(3) $\boldsymbol{g}_i^{\mathrm{T}}\boldsymbol{d}^{(i)}=-\boldsymbol{g}_i^{\mathrm{T}}\boldsymbol{g}_i,(\boldsymbol{d}^{(i)}\neq 0)$.

定理 2.2.11 表明，Fletcher-Reeves 共轭梯度法所产生的搜索方向 $\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)},\cdots,\boldsymbol{d}^{(m)}$ 是 $\boldsymbol{A}$ 共轭的，根据定理 2.2.8，经有限次迭代必达到极小点.

这里要着重指出，选择最速下降方向作初始搜索方向（即 $\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{x}^{(1)})$）是十分重要的. 如果选择别的方向作为初始搜索方向，其余方向均按 Fletcher-Reeves 方法构造，那么极小化正定二次函数时，这样构造出来的一组方向不能保证共轭性.

**例 2.2.12** 考虑问题

$$\min \quad x_1^2+\frac{1}{2}x_2^2+\frac{1}{2}x_3^2,$$

取初始点和初始搜索方向分别为

$$\boldsymbol{x}^{(1)}=\begin{bmatrix}1\\1\\1\end{bmatrix} \text{和}\ \boldsymbol{d}^{(1)}=\begin{bmatrix}-1\\-2\\0\end{bmatrix}.$$

显然，$\boldsymbol{d}^{(1)}$ 不是目标函数在 $\boldsymbol{x}^{(1)}$ 处的最速下降方向. 下面，用 Fletcher-Reeves 法构造两个搜索方向.

首先，从 $\boldsymbol{x}^{(1)}$ 出发，沿方向 $\boldsymbol{d}^{(1)}$ 搜索，求步长 $\lambda_1$，使它满足

$$f(\boldsymbol{x}^{(1)}+\lambda_1\boldsymbol{d}^{(1)})=\min_{\lambda} f(\boldsymbol{x}^{(1)}+\lambda\boldsymbol{d}^{(1)}),$$

得到 $\lambda_1=2/3$，从而得出

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\lambda_1\boldsymbol{d}^{(1)}=\begin{bmatrix}\frac{1}{3}\\-\frac{1}{3}\\1\end{bmatrix},$$

$$\boldsymbol{g}_2=\begin{bmatrix}\frac{2}{3}\\-\frac{1}{3}\\1\end{bmatrix}.$$

令 $\boldsymbol{d}^{(2)}=-\boldsymbol{g}_2+\beta_1\boldsymbol{d}^{(1)}$，根据公式(2.34)，有

$$\beta_1=\frac{\boldsymbol{d}^{(1)\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_2}{\boldsymbol{d}^{(1)\mathrm{T}}\boldsymbol{A}\boldsymbol{d}^{(1)}}=-\frac{1}{9},$$ 则

$$\boldsymbol{d}^{(2)}=\begin{bmatrix}-\frac{5}{9}\\\frac{5}{9}\\-1\end{bmatrix}.$$

再从 $\boldsymbol{x}^{(2)}$ 出发，沿 $\boldsymbol{d}^{(2)}$ 搜索，求得步长 $\lambda_2=21/26$，从而得到

$$\boldsymbol{x}^{(3)}=\boldsymbol{x}^{(2)}+\lambda_2\boldsymbol{d}^{(2)}=\begin{bmatrix}-\frac{9}{78}\\\frac{9}{78}\\\frac{5}{26}\end{bmatrix},$$

$$\boldsymbol{g}_3=\begin{bmatrix}-\frac{18}{78}\\\frac{9}{78}\\\frac{5}{26}\end{bmatrix}.$$

令 $\boldsymbol{d}^{(3)}=-\boldsymbol{g}_3+\beta_2\boldsymbol{d}^{(2)}$，利用公式(2.34)算出 $\beta_2$. 从而得到搜索方向

$$\boldsymbol{d}^{(3)}=\frac{1}{676}\begin{bmatrix}131\\-53\\-175\end{bmatrix}.$$

容易验证，$\boldsymbol{d}^{(1)}$ 与 $\boldsymbol{d}^{(2)}$ 关于 $\boldsymbol{A}$ 共轭，$\boldsymbol{d}^{(2)}$ 与 $\boldsymbol{d}^{(3)}$ 关于 $\boldsymbol{A}$ 共轭，但是 $\boldsymbol{d}^{(1)}$ 与 $\boldsymbol{d}^{(3)}$ 不关于 $\boldsymbol{A}$ 共轭，因此向量组 $\boldsymbol{d}^{(1)},\boldsymbol{d}^{(2)},\boldsymbol{d}^{(3)}$ 不是关于 $\boldsymbol{A}$ 共轭的. 在 Fletcher-Reeves 法中，初始搜索方向必须取最速下降方向，这一点绝不可忽视.

### 3 用于一般函数的共轭梯度法

Fletcher-Reeves 共轭梯度法用于一般可微函数时，迭代公式与(2.34)略有不同，由于 $f(\boldsymbol{x})$ 不一定为二次凸函数，因此在(2.34)中一维搜索步长公式不再适用. 这时，可用 2.1.5 中介绍的一维搜索方法求步长 $\lambda_k$.

$$\lambda_k:\ f(\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)})=\min_{\lambda\geq 0}f(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)}). \tag{2.35}$$

其余公式不变，即可用于一般函数的情形. Fletcher-Reeves 共轭梯度法的计算步骤如下：

(1) 给定点 $\boldsymbol{x}^{(1)}$，允许误差 $\epsilon>0$. 置 $\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(1)}$，$\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{y}^{(1)})$，$k=j=1$.

(2) 若 $\|\nabla f(\boldsymbol{y}^{(j)})\|<\epsilon$，则停止计算；否则，作一维搜索，求 $\lambda_j$，满足

$$f(\boldsymbol{y}^{(j)}+\lambda_j\boldsymbol{d}^{(j)})=\min_{\lambda\geq 0}f(\boldsymbol{y}^{(j)}+\lambda\boldsymbol{d}^{(j)}),$$

令 $\boldsymbol{y}^{(j+1)}=\boldsymbol{y}^{(j)}+\lambda_j\boldsymbol{d}^{(j)}$.

(3) 如果 $j<n$，则作(4)；否则，进行(5).

(4) 计算 $\nabla f(\boldsymbol{y}^{(j+1)})$，置 $\boldsymbol{d}^{(j+1)}=-\nabla f(\boldsymbol{y}^{(j+1)})+\beta_j\boldsymbol{d}^{(j)}$，其中

$$\beta_j=\frac{\|\nabla f(\boldsymbol{y}^{(j+1)})\|^2}{\|\nabla f(\boldsymbol{y}^{(j)})\|^2}. \tag{2.36}$$

置 $j:=j+1$，转(2).

(5) 令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{y}^{(n+1)}$，$\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(k+1)}$，$\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{y}^{(1)})$，置 $j=1$，

$k := k+1$，转(2).

Fletcher-Reeves 法用于求任意函数的极小点时，一般说来，用有限步迭代是达不到的. 迭代的延续可以采取不同的方案. 这里是把 $n$ 步作为一轮，每搜索一轮之后，取一次最速下降方向，开始下一轮. 这种策略称为"重新开始"或"重置". 每 $n$ 次迭代后以最速下降方向重新开始的共轭梯度法，有时称为传统的共轭梯度法.

共轭梯度法中，可以采取不同的公式计算 $\beta_j$(其他不变)，因而形成不同的共轭梯度法. 有以下几种常见形式：

(1) Daniel 法

这种方法中，计算 $\beta_j$ 的公式为

$$\beta_j = \frac{\boldsymbol{d}^{(j)\mathrm{T}} \nabla^2 f(\boldsymbol{x}^{(j+1)}) \boldsymbol{g}_{j+1}}{\boldsymbol{d}^{(j)\mathrm{T}} \nabla^2 f(\boldsymbol{x}^{(j+1)}) \boldsymbol{d}^{(j)}}, \tag{2.37}$$

(2) Sorenson-Wolfe 法

计算 $\beta_j$ 的公式是

$$\beta_j = \frac{\boldsymbol{g}_{j+1}^{\mathrm{T}}(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}{\boldsymbol{d}^{(j)\mathrm{T}}(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}. \tag{2.38}$$

(3) Polak-Ribiere-Polyak 法

计算 $\beta_j$ 的公式为

$$\beta_j = \frac{\boldsymbol{g}_{j+1}^{\mathrm{T}}(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)}{\boldsymbol{g}_j^{\mathrm{T}} \boldsymbol{g}_j}. \tag{2.39}$$

当极小化正定二次函数，初始搜索方向取负梯度时，以上三种方法与 FR 共轭梯度法等价. 但是，对一般函数 $f(\boldsymbol{x})$，上述四种方法(2.34)，(2.37)～(2.39)求得的搜索方向并不相同. 这些共轭梯度法计算效果大体一致，由于 FR 法(2.34)便于记忆和实现，所以人们往往把它作为共轭梯度法的代表.

共轭梯度法的收敛速度介于最速下降法与 Newton 法之间，迭代要周期进行，即每迭代 $n$ 次要重新开始.

**4　Best 加速**

为提高收敛速度，在共轭梯度法中，沿 $n$ 个共轭方向搜索之后，根据计算情况可采取一个加速步骤. 加速方案如下所述：

已知一组共轭方向 $d^{(1)},d^{(2)},\cdots,d^{(n)}$；从 $x^{(1)}$ 出发，沿这组方向搜索得到点列 $x^{(1)},x^{(2)},\cdots,x^{(n+1)}$. 各点处的梯度记作 $g_1,g_2,\cdots,g_{n+1}$.

(1) 记矩阵 $M=(d^{(1)},d^{(2)},\cdots,d^{(n)})$.

(2) 求向量组

$$q_j=\frac{g_{j+1}-g_j}{\|x^{(j+1)}-x^{(j)}\|},\ j=1,2,\cdots,n.$$

构成矩阵 $D=(q_1,q_2,\cdots,q_n)$

(3) 求解 $w$

$$Dw=g_{n+1}.$$

(4) 求 $s=-Mw,\alpha=g_{n+1}^{T}s$.

(5) 若 $\alpha<0$，则求 $x^{(n+2)}=x^{(n+1)}+\lambda_{n+1}s$，其中 $\lambda_{n+1}$ 由一维搜索确定，即使得

$$f(x^{(n+2)})=\min_{\lambda}f(x^{(n+1)}+\lambda s),$$

这样完成一个加速步骤.

若 $\alpha\geqslant0$，则不加速.

以上从 $x^{(n+1)}$ 到 $x^{(n+2)}$ 称为一个**加速步**.

### 2.2.4 拟 Newton 法

前面介绍的 Newton 法，突出优点是收敛很快. 但是，运用 Newton 法需要计算二阶偏导数，而且目标函数的 Hesse 矩阵可能是非正定. 为了克服 Newton 法的缺点，人们提出了拟 Newton 法. 它的基本思想是，用不包含二阶导数的矩阵 $H_k$ 取代 Newton 法中的 Hesse 逆矩阵 $\nabla^2 f(x^{(k)})^{-1}$，再沿方向 $d^{(k)}=-H_kg_k$ 作一维搜索. 由于构造近似矩阵 $H_k$ 的方法不同，因而有不同的拟 Newton 法. 经理论证明和实践检验，拟 Newton 法已经成为一类公认

的比较有效的算法.

1　DFP 法

构造 Hesse 逆矩阵最早的,最巧妙的一种格式,由 Davidon 提出,后来由 Fletcher 和 Powell 作了改进,称为 DFP 法,也称为**变尺度法**(variable metric method).

公式(DFP)

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{p}^{(k)}\boldsymbol{p}^{(k)\mathrm{T}}}{\boldsymbol{p}^{(k)\mathrm{T}}\boldsymbol{q}^{(k)}} - \frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)}}, \tag{2.40}$$

$\boldsymbol{H}_1$ 为给定的 $n$ 阶对称正定矩阵,一般令 $\boldsymbol{H}_1$ 为单位矩阵 $\boldsymbol{I}_n$($n$ 阶单位矩阵).

计算步骤:

(1) 给定初始点 $\boldsymbol{x}^{(1)} \in E^n$,允许误差 $\varepsilon > 0$.

(2) 置 $\boldsymbol{H}_1 = \boldsymbol{I}_n$,计算出在 $\boldsymbol{x}^{(1)}$ 处的梯度 $\boldsymbol{g}_1 = \nabla f(\boldsymbol{x}^{(1)})$. 置 $k = 1$.

(3) 令 $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k\boldsymbol{g}_k$.

(4) 从 $\boldsymbol{x}^{(k)}$ 出发,沿方向 $\boldsymbol{d}^{(k)}$ 搜索,求步长 $\lambda_k$,使它满足

$$f(\boldsymbol{x}^{(k)} + \lambda_k\boldsymbol{d}^{(k)}) = \min_{\lambda \geqslant 0} f(\boldsymbol{x}^{(k)} + \lambda\boldsymbol{d}^{(k)}).$$

令 $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k\boldsymbol{d}^{(k)}$,计算 $\boldsymbol{g}_{k+1} = \nabla f(\boldsymbol{x}^{(k+1)})$.

(5) 检验是否满足收敛准则,若满足 $\| \boldsymbol{g}_{k+1} \| \leqslant \varepsilon$,则停止迭代,得到 $\bar{\boldsymbol{x}} = \boldsymbol{x}^{(k+1)}$. 否则,进行(6).

(6) 若 $k = n$,则令 $\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(k+1)}$,返回(2);否则,进行(7).

(7) 令 $\boldsymbol{p}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$,$\boldsymbol{q}^{(k)} = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$. 利用公式(2.40)计算 $\boldsymbol{H}_{k+1}$,置 $k := k+1$,返回(3).

**例 2.2.13**　用 DFP 方法求解下列问题:

$$\min \quad 2x_1^2 + x_2^2 - 4x_1 + 2,$$

初始点及初始矩阵分别取为

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \boldsymbol{H}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

**解** 第1次迭代：

在 $\boldsymbol{x}^{(1)}$ 处的梯度及搜索方向为

$$\boldsymbol{g}_1=\begin{bmatrix}4\\2\end{bmatrix},\boldsymbol{d}^{(1)}=-\boldsymbol{H}_1\boldsymbol{g}_1=\begin{bmatrix}-4\\-2\end{bmatrix}.$$

从 $\boldsymbol{x}^{(1)}$ 出发，沿方向 $\boldsymbol{d}^{(1)}$ 作一维搜索，求步长 $\lambda_1$：

$$\min_{\lambda\geq 0}\ f(\boldsymbol{x}^{(1)}+\lambda\boldsymbol{d}^{(1)}),$$

得到步长 $\lambda_1=5/18$. 因此沿 $\boldsymbol{d}^{(1)}$ 方向搜索得到的点 $\boldsymbol{x}^{(2)}$ 及在此点目标函数的梯度为

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\lambda_1\boldsymbol{d}^{(1)}=\begin{bmatrix}\frac{8}{9}\\\frac{4}{9}\end{bmatrix},$$

$$\boldsymbol{g}_2=\begin{bmatrix}-\frac{4}{9}\\\frac{8}{9}\end{bmatrix}.$$

第2次迭代：

令

$$\boldsymbol{p}^{(1)}=\lambda_1\boldsymbol{d}^{(1)}=\begin{bmatrix}-\frac{10}{9}\\-\frac{5}{9}\end{bmatrix},$$

$$\boldsymbol{q}^{(1)}=\boldsymbol{g}_2-\boldsymbol{g}_1=\begin{bmatrix}-\frac{40}{9}\\-\frac{10}{9}\end{bmatrix}.$$

计算矩阵 $\boldsymbol{H}_2$ 及搜索方向 $\boldsymbol{d}^{(2)}$：

$$\boldsymbol{H}_2=\boldsymbol{H}_1+\frac{\boldsymbol{p}^{(1)}\boldsymbol{p}^{(1)\mathrm{T}}}{\boldsymbol{p}^{(1)\mathrm{T}}\boldsymbol{q}^{(1)}}-\frac{\boldsymbol{H}_1\boldsymbol{q}^{(1)}\boldsymbol{q}^{(1)\mathrm{T}}\boldsymbol{H}_1}{\boldsymbol{q}^{(1)\mathrm{T}}\boldsymbol{H}_1\boldsymbol{q}^{(1)}}$$

$$=\frac{1}{306}\begin{bmatrix} 86 & -38 \\ -38 & 305 \end{bmatrix}.$$

$$d^{(2)}=-H_2g_2=\frac{12}{51}\begin{bmatrix} 1 \\ -4 \end{bmatrix}.$$

从 $x^{(2)}$ 出发,沿方向 $d^{(2)}$ 搜索:

$$\min_{\lambda\geqslant 0} f(x^{(2)}+\lambda d^{(2)}),$$

得到 $\lambda_2=17/36$,经第 2 次迭代得到的点及在此点目标函数的梯度分别为

$$x^{(3)}=x^{(2)}+\lambda_2 d^{(2)}=\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{及 } g_3=\nabla f(x^{(3)})=\begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

因此,$x^{(3)}$ 为最优解.

DFP 算法具有正定性及二次终止性.

**定理 2.2.14** 若 $g_i\neq 0, i=1,2,\cdots,n$,则用 DFP 方法构造的矩阵 $H_i(i=1,2,\cdots,n)$ 为对称正定矩阵.

根据这个定理,DFP 方法中搜索方向

$$d^{(k)}=-H_kg_k \tag{2.41}$$

必为下降方向.

**定理 2.2.15** 设用 DFP 方法解下列问题:

$$\min \quad f(x)\triangleq\frac{1}{2}x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c,$$

其中 $A$ 为 $n$ 阶对称正定矩阵. 初始点 $x^{(1)}\in E^n$,令 $H_1$ 为 $n$ 阶对称正定矩阵,则成立

(1) $p^{(i)\mathrm{T}}Ap^{(j)}=0,\ 1\leqslant i<j\leqslant k$,

(2) $H_{k+1}Ap^{(i)}=p^{(i)}, 1\leqslant i\leqslant k$.

其中 $p^{(i)}=x^{(i+1)}-x^{(i)}=\lambda_i d^{(i)}, \lambda_i\neq 0, k\leqslant n$.

上述定理 2.2.15 中,当 $k=n$ 时,有 $H_{n+1}=A^{-1}$.

它表明,DFP 方法中构造出来的搜索方向是一组 $A$ 共轭方向,因此 DFP 方法具有二次终止性.

关于 DFP 方法用于一般函数时的收敛性，有如下结论：

如果 $f$ 是 $E^n$ 上的二次连续可微实函数，对任意的 $\hat{x}\in E^n$，存在常数 $m>0$，使得当 $x\in C(\hat{x})=\{x\mid f(x)\leqslant f(\hat{x})\}$，$y\in E^n$ 时，有

$$m\|y\|^2\leqslant y^{\mathrm{T}}\nabla^2 f(x)y, \tag{2.42}$$

则用 DFP 方法产生的序列 $\{x^{(k)}\}$ 或终止于或收敛于 $f$ 在 $E^n$ 上的唯一极小点.

## 2 拟 Newton 法的其他形式

公式 BFGS

$$\begin{aligned} H_{k+1}^{\mathrm{BFGS}} = H_k &+ \left(1+\frac{q^{(k)\mathrm{T}}H_k q^{(k)}}{p^{(k)\mathrm{T}}q^{(k)}}\right)\frac{p^{(k)}p^{(k)\mathrm{T}}}{p^{(k)\mathrm{T}}q^{(k)}} \\ &- \frac{p^{(k)}q^{(k)\mathrm{T}}H_k + H_k q^{(k)}p^{(k)\mathrm{T}}}{p^{(k)\mathrm{T}}q^{(k)}}. \end{aligned} \tag{2.43}$$

这个重要公式是由 Broyden，Fletcher，Goldfarb 和 Shanno 于 1970 年提出的. 它可以像 DFP 公式(2.40)一样使用. 数值计算表明，BFGS 算法比较好，目前得到广泛应用. 计算步骤与 DFP 法相同，只需将 $H_k$ 改为 $H_k^{\mathrm{BFGS}}$.

公式(秩 1 校正)

$$H_{k+1} = H_k + \frac{(p^{(k)}-H_k q^{(k)})(p^{(k)}-H_k q^{(k)})^{\mathrm{T}}}{q^{(k)\mathrm{T}}(p^{(k)}-H_k q^{(k)})} \tag{2.44}$$

计算步骤见 DFP 方法，其中凡计算 $H_k$ 之处，均需改用公式(2.44).

采用秩 1 校正(rank-one correction)公式，算法在一定条件下是收敛的，并具有二次终止性. 然而也存在一些困难. 首先，仅当

$$q^{(k)\mathrm{T}}(p^{(k)}-H_k q^{(k)})>0 \tag{2.45}$$

时，由(2.44)得到的 $H_{k+1}$ 才能保证正定性；其次，即使(2.45)成立，由于舍入误差的影响，可能产生数值计算上的困难.

## 3 自选尺度的拟 Newton 算法

前面介绍的几种变尺度法，具有二次终止性. 这表示算法在早

期阶段有迅速进行的可能性. 对于 $n$ 很大的问题,往往希望在完成 $n$ 次迭代之前就结束下降过程,因此保证算法在每一阶段都能迅速进行才是本质问题. 自选尺度算法恰在这一点上具有优越性,它在每一步具有有利的特征值结构.

计算步骤:

(1) 给定初始点 $\boldsymbol{x}^{(1)}\in E^n$,允许误差 $\varepsilon>0$.

(2) 令 $\boldsymbol{H}_1=\boldsymbol{I}_n$, $\boldsymbol{g}_1=\nabla f(\boldsymbol{x}^{(1)})$,置 $k=1$.

(3) 令 $\boldsymbol{d}^{(k)}=-\boldsymbol{H}_k\boldsymbol{g}_k$.

(4) 从 $\boldsymbol{x}^{(k)}$ 出发,沿方向 $\boldsymbol{d}^{(k)}$ 搜索,求步长 $\lambda_k$,使它满足:

$$f(\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)})=\min_{\lambda\geqslant 0}f(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)})$$

令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$, $\boldsymbol{g}_{k+1}=\nabla f(\boldsymbol{x}^{(k+1)})$.

(5) 若 $\|\boldsymbol{g}_{k+1}\|\leqslant\varepsilon$,则停止迭代,得到点 $\bar{\boldsymbol{x}}=\boldsymbol{x}^{(k+1)}$;否则,进行(6).

(6) 若 $k=n$,则令 $\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(k+1)}$,返回(2);否则,进行(7).

(7) 令 $\boldsymbol{p}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}$, $\boldsymbol{q}^{(k)}=\boldsymbol{g}_{k+1}-\boldsymbol{g}_k$

$$\boldsymbol{H}_{k+1}=\left(\boldsymbol{H}_k-\frac{\boldsymbol{H}_k\boldsymbol{q}^{(k)}\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k}{\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)}}\right)\frac{\boldsymbol{p}^{(k)\mathrm{T}}\boldsymbol{q}^{(k)}}{\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)}}+\frac{\boldsymbol{p}^{(k)}\boldsymbol{p}^{(k)\mathrm{T}}}{\boldsymbol{p}^{(k)\mathrm{T}}\boldsymbol{q}^{(k)}}$$

置 $k:=k+1$,返回(3).

计算中,尺度因子取作

$$\gamma_k=\frac{\boldsymbol{p}^{(k)\mathrm{T}}\boldsymbol{q}^{(k)}}{\boldsymbol{q}^{(k)\mathrm{T}}\boldsymbol{H}_k\boldsymbol{q}^{(k)}}. \tag{2.46}$$

### 4 适时方法

以最速下降法和 Newton 法的组合作为拟 Newton 法改进的基础,可以给出适时方法,其计算步骤如下:

(1) 给定 $\boldsymbol{x}^{(0)}$,置 $k=0$.

(2) 令 $\boldsymbol{D}_k=\boldsymbol{0}$($n$ 阶矩阵).

(3) 令 $\boldsymbol{d}^{(k)}=-\boldsymbol{D}_k\nabla f(\boldsymbol{x}^{(k)})$.

(4) 求出 $\boldsymbol{z}^{(k)}=\boldsymbol{x}^{(k)}+\beta_k\boldsymbol{d}^{(k)}$,

$$\beta_k: f(\boldsymbol{x}^{(k)}+\beta_k\boldsymbol{d}^{(k)})=\min_{\beta} f(\boldsymbol{x}^{(k)}+\beta\boldsymbol{d}^{(k)}).$$

(5) 求 $\boldsymbol{x}^{(k+1)}=\boldsymbol{z}^{(k)}-\alpha_k\nabla f(\boldsymbol{z}^{(k)})$,

$$\alpha_k: f(\boldsymbol{x}^{(k)}-\alpha_k\nabla f(\boldsymbol{z}^{(k)}))=\min_{\alpha} f(\boldsymbol{x}^{(k)}-\alpha\nabla f(\boldsymbol{z}^{(k)})).$$

令 $\boldsymbol{p}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{z}^{(k)}$,

$$\boldsymbol{q}^{(k)}=\nabla f(\boldsymbol{x}^{(k+1)})-\nabla f(\boldsymbol{z}^{(k)}).$$

(6) 如果 $j=n$,则令 $j=0$,并置 $k:=k+1$,返回(2);否则,进行(7).

(7) 如果$(\boldsymbol{p}^{(k)}-\boldsymbol{D}_k\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)}>0$,令

$$\boldsymbol{D}_{k+1}=\boldsymbol{D}_k+\frac{(\boldsymbol{p}^{(k)}-\boldsymbol{D}_k\boldsymbol{q}^{(k)})(\boldsymbol{p}^{(k)}-\boldsymbol{D}_k\boldsymbol{q}^{(k)})^{\mathrm{T}}}{(\boldsymbol{p}^{(k)}-\boldsymbol{D}_k\boldsymbol{q}^{(k)})^{\mathrm{T}}\boldsymbol{q}^{(k)}}$$

置 $j:=j+1$;否则,令

$$\boldsymbol{D}_{k+1}=\boldsymbol{D}_k,$$

置 $k:=k+1$,返回(3).

上述方法具有良好的收敛性,能够保证它每一步至少像最速下降法那样快.

### 2.2.5 最小二乘法

无约束最优化问题中,有些重要的特殊情形,比如目标函数由若干个函数的平方和构成.这类函数一般可以写成

$$F(\boldsymbol{x})=\sum_{i=1}^{m}f_i^2(\boldsymbol{x}),\quad x\in E^n \tag{2.47}$$

其中 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$,一般假设 $m\geqslant n$.我们把极小化这类函数的问题

$$\min\quad F(\boldsymbol{x})\triangleq\sum_{i=1}^{m}f_i^2(\boldsymbol{x}) \tag{2.48}$$

称为**最小二乘问题**(least square problem).当每个 $f_i(\boldsymbol{x})$为 $\boldsymbol{x}$ 的线

性函数时，称为**线性最小二乘问题**. 当 $f_i(x)$是 $x$ 的非线性函数时，称为**非线性最小二乘问题**.

由于目标函数 $F(x)$具有若干个函数平方和这种特殊形式，因此给求解带来方便. 这类问题，除了能够运用一般求解方法外，还有更为简便有效的解法. 下面介绍**最小二乘法**(least square method).

**1 线性最小二乘问题的求解公式**

假设在(2.47)中，$f_i(x)$为线性函数：

$$f_i(\boldsymbol{x}) = \boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{x} - b_i, i = 1,\cdots,m.$$

其中 $\boldsymbol{P}_i$ 是 $n$ 维列向量，$b_i$ 是实数. 则可以用矩阵乘积形式来表达

$$\sum_{i=1}^{m} f_i^2(\boldsymbol{x}).$$

记作
$$\boldsymbol{A} \triangleq \begin{bmatrix} \boldsymbol{p}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{p}_m^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{b} \triangleq \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix},$$

$\boldsymbol{A}$ 是 $m\times n$ 矩阵，$\boldsymbol{b}$ 是 $m$ 维列向量. 则

$$F(\boldsymbol{x}) = \sum_{i=1}^{m} f_i^2(\boldsymbol{x})$$
$$= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}).$$

设 $\boldsymbol{A}$ 列满秩，$\mathbf{A}^{\mathrm{T}}\boldsymbol{A}$ 为 $n$ 阶对称正定矩阵. 目标函数 $F(\boldsymbol{x})$的整体极小点为

$$\bar{\boldsymbol{x}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \tag{2.49}$$

对于线性最小二乘问题，只要 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 非奇异，就可用公式(2.49)求解.

下面给出一个用最小二乘法求解线性方程组的例. 具体求解方法是，用公式给出最小二乘解，再分析该解是否为线性方程组的解.

**例 2.2.16** 求解下列线性方程组

$$\begin{cases} x_1 + x_2 = 3, \\ 2x_1 - 3x_2 = 2, \\ -x_1 + 4x_2 = 4. \end{cases}$$

**解** 设方程组的系数矩阵为 $\boldsymbol{A}$,右端向量为 $\boldsymbol{b}$:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 2 & -3 \\ -1 & 4 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix},$$

将方程组写作 $\boldsymbol{Ax}=\boldsymbol{b}$.

求二次函数 $f(\boldsymbol{x})=(\boldsymbol{Ax}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax}-\mathrm{b})$的极小点.

先计算$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$:

$$A^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 6 & -9 \\ -9 & 26 \end{bmatrix}, \quad (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1} = \begin{bmatrix} \frac{26}{75} & \frac{9}{75} \\ \frac{9}{75} & \frac{6}{75} \end{bmatrix}.$$

再根据公式(2.49),算出 $f(\boldsymbol{x})$的极小点

$$\bar{\boldsymbol{x}} = \begin{bmatrix} \frac{13}{5} \\ \frac{7}{5} \end{bmatrix},$$

这个极小点称为最小二乘解.

函数 $f(\boldsymbol{x})$的极小值

$$f_{\min} = (\boldsymbol{A}\bar{\boldsymbol{x}} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\bar{\boldsymbol{x}} - \boldsymbol{b}) = 3.$$

此例中,$f_{\min}\neq 0$ 表明,线性方程组无解.当方程组有解时,最小二乘解也是线性方程组的解.

**2 非线性最小二乘法**

设在(2.47)中 $f_i(\boldsymbol{x})$是非线性函数,且 $F(\boldsymbol{x})$存在连续偏导数.由于 $f_i(\boldsymbol{x})$是非线性函数,(2.48)为非线性最小二乘问题,因此不能使用公式(2.49).解非线性最小二乘问题的基本思想是,通过解一系列线性最小二乘问题求非线性最小二乘问题的解.设 $\boldsymbol{x}^{(k)}$是

解的第 $k$ 次近似. 在 $\boldsymbol{x}^{(k)}$将 $f_i(\boldsymbol{x})$线性化,把原来问题转化成线性最小二乘问题,运用公式(2.49)求出此问题的极小点 $\boldsymbol{x}^{(k+1)}$,把它作为非线性最小二乘问题解的第 $k+1$ 次近似. 再从 $\boldsymbol{x}^{(k+1)}$出发,重复以上过程.

迭代公式如下:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{f}^{(k)}. \tag{2.50}$$

其中

$$\boldsymbol{A}_k = \begin{bmatrix} \nabla f_1(\boldsymbol{x}^{(k)})^{\mathrm{T}} \\ \vdots \\ \nabla f_m(\boldsymbol{x}^{(k)})^{\mathrm{T}} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x}^{(k)})}{\partial x_1}, \dfrac{\partial f_1(\boldsymbol{x}^{(k)})}{\partial x_2}, \cdots, \dfrac{\partial f_1(\boldsymbol{x}^{(k)})}{\partial x_n} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial f_m(\boldsymbol{x}^{(k)})}{\partial x_1}, \dfrac{\partial f_m(\boldsymbol{x}^{(k)})}{\partial x_2}, \cdots, \dfrac{\partial f_m(\boldsymbol{x}^{(k)})}{\partial x_n} \end{bmatrix},$$

$$\boldsymbol{f}^{(k)} = \begin{bmatrix} f_1(\boldsymbol{x}^{(k)}) \\ f_2(\boldsymbol{x}^{(k)}) \\ \vdots \\ f_m(\boldsymbol{x}^{(k)}) \end{bmatrix}.$$

$\boldsymbol{x}^{(k+1)}$作为 $F(\boldsymbol{x})$的极小点的第 $k+1$ 次近似.

公式(2.50)可以写作

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{H}_k^{-1}\nabla F(\boldsymbol{x}^{(k)}) \tag{2.51}$$

其中

$$\nabla F(\boldsymbol{x}^{(k)}) = 2\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{f}^{(k)} \tag{2.52}$$

为目标函数 $F(\boldsymbol{x})$在点 $\boldsymbol{x}^{(k)}$处的梯度.

$\boldsymbol{H}_k^{-1}$ 为 $\boldsymbol{H}_k$ 的逆阵.

$$\boldsymbol{H}_k = 2\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k. \tag{2.53}$$

显然,(2.51)与 Newton 迭代公式类似. 通常称(2.50)或

(2.51)为高斯-牛顿公式.

**定义 2.2.17** 向量 $\boldsymbol{d}^{(k)}=-(\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{f}^{(k)}$称为在点 $\boldsymbol{x}^{(k)}$处的高斯-牛顿方向.

为保证每次迭代能使目标函数值下降(至少不能上升),在求出方向 $\boldsymbol{d}^{(k)}$后,不直接用 $\boldsymbol{x}^{(k)}+\boldsymbol{d}^{(k)}$作为第 $k+1$ 次近似,而是从 $\boldsymbol{x}^{(k)}$出发,沿这个方向进行一维搜索:

$$\min_{\lambda} F(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)}),$$

求出步长 $\lambda_k$ 后,令

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)},$$

把 $\boldsymbol{x}^{(k+1)}$作为第 $k+1$ 次近似.依此类推,直至得到满足要求的解.

计算步骤:

(1) 给定初点 $\boldsymbol{x}^{(1)}$,允许误差 $\varepsilon>0$,置 $k=1$.

(2) 计算函数值 $f_i(\boldsymbol{x}^{(k)}),i=1,\cdots,m$,得到公式(2.50)中的向量 $\boldsymbol{f}^{(k)}$.再计算一阶偏导数

$$a_{ij}=\frac{\partial f_i(\boldsymbol{x}^{(k)})}{\partial x_j},i=1,\cdots,m;j=1,\cdots,n.$$

得到 $m\times n$ 矩阵 $\boldsymbol{A}_k=(a_{ij})_{m\times n}$.

(3) 解方程组

$$\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{d}=-\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{f}^{(k)}, \tag{2.54}$$

求得高斯-牛顿方向 $\boldsymbol{d}^{(k)}$.

(4) 从 $\boldsymbol{x}^{(k)}$出发,沿 $\boldsymbol{d}^{(k)}$作一维搜索.求步长 $\lambda_k$,使得

$$F(\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)})=\min_{\lambda} F(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)}),$$

令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}$.

(5) 若 $\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\|\leqslant\varepsilon$,则停止计算,得解 $\bar{\boldsymbol{x}}=\boldsymbol{x}^{(k+1)}$;否则,置 $k:=k+1$,返回(2).

### 3 最小二乘法的改进

前面介绍的最小二乘法,有时会出现矩阵 $\boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{A}_k$ 是奇异阵或接

近奇异阵. 这时求$(A_k^TA_k)^{-1}$或解方程组(2.54)会遇到很大困难，甚至不能进行. 因此，人们对最小二乘法作了进一步修正. 所用的基本技巧是把一个正定对角矩阵加到$A_k^TA_k$上去，改变原矩阵的特征值结构，使它变成条件数较好的对称正定矩阵，从而给出**修正的最小二乘法**.

下面给出修正方法之一，Marquardt 法. 在这种算法中，令

$$d^{(k)} = -(A_k^TA_k + \alpha_k I)^{-1}A_k^T f^{(k)} \tag{2.55}$$

其中$I$为$n$阶单位矩阵，$\alpha_k$是一个正实数. 显然，当$\alpha_k=0$时，$d^{(k)}$就是 Gauss-Newton 方向. 当$\alpha_k$充分大时，逆矩阵$(A_k^TA_k+\alpha_k I)^{-1}$主要取决于$\alpha_k I$，这时$d^{(k)}$接近$F(x)$在点$x^{(k)}$处的最速下降方向$(-\nabla F(x^{(k)}))$. 一般地，当$\alpha_k\in(0,+\infty)$时，(2.55)所确定的方向$d^{(k)}$介于 Gauss-Newton 方向与最速下降方向之间. 实际计算中，$\alpha_k$需选用合适的数值，不宜过大，也不宜太小.

计算步骤：

(1) 给定初始点$x^{(1)}$，初始参数$\alpha_1>0$，增长因子$\beta>1$，允许误差$\epsilon>0$，计算$F(x^{(1)})$，置$\alpha=\alpha_1$，$k=1$.

(2) 置$\alpha := \alpha/\beta$，计算

$$f^{(k)} = (f_1(x^{(k)}),\cdots,f_m(x^{(k)}))^T,$$

$$A_k = \begin{bmatrix} \frac{\partial f_1(x^{(k)})}{\partial x_1},\cdots,\frac{\partial f_1(x^{(k)})}{\partial x_n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial f_m(x^{(k)})}{\partial x_1},\cdots,\frac{\partial f_m(x^{(k)})}{\partial x_n} \end{bmatrix}.$$

(3) 解方程

$$(A_k^TA_k + \alpha I)d = -A_k^T f^{(k)},$$

求得方向$d^{(k)}$，令

$$x^{(k+1)} = x^{(k)} + d^{(k)}.$$

(4) 计算$F(x^{(k+1)})$. 若$F(x^{(k+1)})<F(x^{(k)})$，则转步(6)；否则，

进行(5).

(5) 若 $\| A_k^T f^{(k)} \| \leqslant \varepsilon$,则停止计算,得到解 $\bar{x}=x^{(k)}$;否则,置 $\alpha := \beta\alpha$,转(3).

(6) 若 $\| A_k^T f^{(k)} \| \leqslant \varepsilon$,则停止计算,得到解 $\bar{x}=x^{(k+1)}$;否则,置 $k := k+1$,返回(2).

初始参数 $\alpha_1$ 和因子 $\beta$ 应取适当数值,比如,根据经验可取 $\alpha_1 = 0.01$, $\beta=10$.

在上述算法中,若把停步准则改为当梯度 $\nabla F(x^{(k)})=0$ 时算法终止,则 Marquardt 法可能产生无穷序列 $\{x^{(k)}\}$. 这样,关于算法的收敛性,有下列定理:

**定理 2.2.18** 设 $\hat{x}\in E^n$, $F(\hat{x})=\sigma$,水平集

$$S_\sigma = \{x \mid F(x) \leqslant \sigma\}$$

有界,由(2.53)所定义的 $H_k$ 在 $S_\sigma$ 上恒为正定矩阵,初始点 $x^{(1)}\in S_\sigma$,则按 Marquardt 法产生的序列 $\{x^{(k)}\}$满足.

(1) 当 $\{x^{(k)}\}$ 为有穷序列时,序列的最后一个元素是 $F(x)$ 的稳定点;

(2) 当 $\{x^{(k)}\}$ 为无穷序列时,它必有极限点,而且极限点必为 $F(x)$ 的稳定点.

### 2.2.6 模式搜索法

**模式搜索法**(pattern search method)是由 Hooke 和 Jeeves 于 1961 年提出的,因此又称为 Hooke-Jeeves 方法. 其基本思想,从几何上讲,是寻找具有较小函数值的"山谷",力图使迭代产生的序列沿"山谷"逼近极小点. 算法从初始基点开始,包括两种类型的移动,这就是**探测移动**(exploratory move)和**模式移动**(pattern move). 探测移动依次沿 $n$ 个坐标轴进行,用以确定新的基点和有利于函数值下降的方向. 模式移动沿相邻两个基点连线方向进行,试图顺着"山谷"使函数值更快减少. 两种移动交替进行.

设目标函数为 $f(\boldsymbol{x}),\boldsymbol{x}\in E^n$，第 $j$ 个坐标方向记作

$$\boldsymbol{e}_j=(0,\cdots,0,\underset{j}{1},0,\cdots,0)^{\mathrm{T}}.$$

计算步骤：

(1) 给定初始点(第一个基点)$\boldsymbol{x}^{(1)}\in E^n$，$n$ 个坐标方向 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$，初始步长 $\delta$，加速因子 $\alpha\geqslant 1$，缩减率 $\beta\in(0,1)$，允许误差 $\varepsilon>0$，置 $\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(1)}$，$k=1$，$j=1$.

(2) 如果 $f(\boldsymbol{y}^{(j)}+\delta\boldsymbol{e}_j)<f(\boldsymbol{y}^{(j)})$，则令 $\boldsymbol{y}^{(j+1)}=\boldsymbol{y}^{(j)}+\delta\boldsymbol{e}_j$，进行(4)；否则，进行(3).

(3) 如果 $f(\boldsymbol{y}^{(j)}-\delta\boldsymbol{e}_j)<f(\boldsymbol{y}^{(j)})$，则令 $\boldsymbol{y}^{(j+1)}=\boldsymbol{y}^{(j)}-\delta\boldsymbol{e}_j$，进行(4)；否则，令 $\boldsymbol{y}^{(j+1)}=\boldsymbol{y}^{(j)}$，进行(4).

(4) 如果 $j<n$，则置 $j:=j+1$，转步(2)；否则，进行(5).

(5) 如果 $f(\boldsymbol{y}^{(n+1)})<f(\boldsymbol{x}^{(k)})$，则进行步(6)；否则，进行(7).

(6) 置基点 $\boldsymbol{x}^{(k+1)}=\boldsymbol{y}^{(n+1)}$，令

$$\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(k+1)}+\alpha(\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}),$$

置 $k:=k+1$，$j=1$，转(2).

(7) 如果 $\delta\leqslant\varepsilon$，则停止迭代，得到点 $\boldsymbol{x}^{(k)}$；否则，置 $\delta:=\beta\delta$，$\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(k)}$，$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}$，$k:=k+1$，$j=1$，转(2).

**例 2.2.19** 用模式搜索法求解下列问题

$$\min\quad f(\boldsymbol{x})\triangleq(1-x_1)^2+5(x_2-x_1^2)^2$$

取初始点 $\boldsymbol{x}^{(1)}$ 及坐标方向 $\boldsymbol{e}_1,\boldsymbol{e}_2$ 分别为

$$\boldsymbol{x}^{(1)}=\begin{bmatrix}2\\0\end{bmatrix},\boldsymbol{e}_1=\begin{bmatrix}1\\0\end{bmatrix},\boldsymbol{e}_2=\begin{bmatrix}0\\1\end{bmatrix}.$$

令 $\delta=1/2$，$\alpha=1$，$\beta=1/2$.

**解** 先在 $\boldsymbol{x}^{(1)}$ 周围进行探测移动. 第 1 轮探测完成后，得到第 2 个基点

$$x^{(2)}=\begin{bmatrix}\frac{3}{2}\\ \frac{1}{2}\end{bmatrix}.$$

再沿方向 $x^{(2)}-x^{(1)}$进行模式移动，将得到的点记作 $y^{(1)}$，则

$$y^{(1)}=x^{(2)}+\alpha(x^{(2)}-x^{(1)})=\begin{bmatrix}1\\1\end{bmatrix}.$$

模式移动后，立即从得到的点 $y^{(1)}$出发，进行第 2 轮探测移动. 这轮探测结果仍然得到点$(1,1)^{\mathrm{T}}$，而在这一点处目标函数值比前一个基点 $x^{(2)}$处的目标函数值有所下降，即

$$f(1,1)=0<f\left(\frac{3}{2},\frac{1}{2}\right)=\frac{249}{16},$$

这表明模式移动是成功的. 因此得到第 3 个基点 $x^{(3)}=(1,1)^{\mathrm{T}}$.

现在进行第 2 次模式移动，这一次从新的基点 $x^{(3)}$出发，沿基点 $x^{(2)}$到 $x^{(3)}$的连线方向进行. 把得到的点仍记作 $y^{(1)}$，

令 $$y^{(1)}=x^{(3)}+\alpha(x^{(3)}-x^{(2)})=\begin{bmatrix}\frac{1}{2}\\ \frac{3}{2}\end{bmatrix}.$$

然后，从 $y^{(1)}$出发，进行第 3 轮探测移动. 作下去将会发现，第 2 次模式移动并不成功. 因此退回到基点 $x^{(3)}$. 减小步长，令 $\delta:=\beta\delta=1/4$. 在 $x^{(3)}$周围进行探测，结果仍失败，必须继续缩减步长. 作下去必能得出结论，$x^{(3)}$是局部最优解. 对此问题，容易用解析方法验证所得到的结论.

在上面例中，探测移动沿各坐标方向所取步长相同. 实际上，不同坐标方向可以给定不同的步长，方法上可有一定的灵活性.

### 2.2.7 Rosenbrock 算法

Rosenbrock 算法又称为转轴法. 它与 Hooke-Jeeves 方法有

类似之处，也是顺着"山谷"求函数的极小点.

Rosenbrock 算法包括探测阶段和构造方向两部分内容. 探测阶段中，从一点出发，依次沿 $n$ 个单位正交方向进行探测移动，一轮探测之后，再从第 1 个方向开始继续探测. 经过若干轮探测移动，完成一个探测阶段. 然后，构造一组新的单位正交方向，称之为转轴，在下一次迭代中，将沿这些方向进行探测.

探测阶段中所用探测方法与 Hooke-Jeeves 方法类似. 构造新的单位正交方向的方法分作两步. 先利用当前的搜索方向和迭代中得到的数据构造一组线性无关的方向，然后利用 Gram-Schmidt 正交化方法，将其正交化及单位化. 具体方法如下：

设原来的搜索方向为

$$\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n)},$$

现在定义一组新的方向

$$\boldsymbol{p}^{(1)}, \boldsymbol{p}^{(2)}, \cdots, \boldsymbol{p}^{(n)},$$

使得

$$\boldsymbol{p}^{(j)} = \begin{cases} \boldsymbol{d}^{(j)}, & \text{当 } \lambda_j = 0; \\ \sum_{i=j}^{n} \lambda_i \boldsymbol{d}^{(i)}, & \text{当 } \lambda_j \neq 0. \end{cases} \tag{2.56}$$

其中 $\lambda_j$ 是探测阶段中沿 $\boldsymbol{d}^{(j)}$ 方向的步长.

然后正交化，令

$$\boldsymbol{q}^{(j)} = \begin{cases} \boldsymbol{p}^{(j)}, & j = 1; \\ \boldsymbol{p}^{(j)} - \sum_{i=1}^{j-1} \dfrac{\boldsymbol{q}^{(i)\mathrm{T}} \boldsymbol{p}^{(j)}}{\boldsymbol{q}^{(i)\mathrm{T}} \boldsymbol{q}^{(i)}} \boldsymbol{q}^{(i)}, & j \geqslant 2. \end{cases} \tag{2.57}$$

再单位化，令

$$\bar{\boldsymbol{d}}^{(j)} = \frac{\boldsymbol{q}^{(j)}}{\| \boldsymbol{q}^{(j)} \|}, j = 1, 2, \cdots, n. \tag{2.58}$$

下一个探测阶段，将沿方向 $\bar{\boldsymbol{d}}^{(j)}$ 进行搜索.

计算步骤如下：

(1) 给定初始点 $x^{(1)}\in E^n$,单位正交方向 $d^{(1)},d^{(2)},\cdots,d^{(n)}$,一般取坐标方向,沿各方向的步长分别为 $\delta_1^{(0)},\delta_2^{(0)},\cdots,\delta_n^{(0)}$,放大因子 $\alpha>1$,缩减因子 $\beta\in(-1,0)$,允许误差 $\varepsilon>0$. 置 $y^{(1)}=x^{(1)},\delta_i=\delta_i^{(0)}$, $i=1,2,\cdots,n$.
$k=1,j=1$.

(2) 如果 $f(y^{(j)}+\delta_j d^{(j)})<f(y^{(j)})$,则令 $y^{(j+1)}=y^{(j)}+\delta_j d^{(j)}$, $\delta_j:=\alpha\delta_j$;

如果 $f(y^{(j)}+\delta_j d^{(j)})\geqslant f(y^{(j)})$,则令 $y^{(j+1)}=y^{(j)},\delta_j:=\beta\delta_j$.

(3) 如果 $j<n$,则置 $j:=j+1$,转(2);否则,进行(4).

(4) 如果 $f(y^{(n+1)})<f(y^{(1)})$,则令 $y^{(1)}=y^{(n+1)}$,置 $j=1$,转(2);如果 $f(y^{(n+1)})=f(y^{(1)})$,则进行(5).

(5) 如果 $f(y^{(n+1)})<f(x^{(k)})$,则进行(6);否则,如果对每个 $j$,成立 $|\delta_j|\leqslant\varepsilon$,则停止计算,$x^{(k)}$作为最优解的估计. 如果不满足终止准则,令 $y^{(1)}=y^{(n+1)}$,置 $j=1$,转(2).

(6) 令 $x^{(k+1)}=y^{(n+1)}$. 如果 $\|x^{(k+1)}-x^{(k)}\|\leqslant\varepsilon$,则取 $x^{(k+1)}$作为极小点的估计,停止计算;否则,计算 $\lambda_1,\lambda_2,\cdots,\lambda_n$,利用公式(2.56)至(2.58)构造新的正交方向 $\bar{d}^{(1)},\bar{d}^{(2)},\cdots,\bar{d}^{(n)}$,并令

$$d^{(j)}=\bar{d}^{(j)},j=1,2,\cdots,n.$$

置 $\delta_j=\delta_j^{(0)},j=1,2,\cdots,n$;置 $y^{(1)}=x^{(k+1)},k:=k+1,j=1$,返回步(2).

由上述计算过程易知,当某轮探测沿 $n$ 个方向均失败,即函数值均不下降时,便完成一个探测阶段. 这时,按算法规定,应进行转轴. 然后,沿着新定义的方向,开始下一个探测阶段.

**例 2.2.20** 用 Rosenbrock 方法解下列问题

$$\min\quad f(x)\triangleq(x_1-3)^2+2(x_2+2)^2.$$

取初始点及初始搜索方向为

$$x^{(1)}=\begin{bmatrix}0\\0\end{bmatrix},d^{(1)}=\begin{bmatrix}1\\0\end{bmatrix},d^{(2)}=\begin{bmatrix}0\\1\end{bmatrix}.$$

初始步长 $\delta_1^{(0)}=\delta_2^{(0)}=1, \alpha=3, \beta=-0.5$.

下面简述求解过程.

先进行第 1 轮探测：

初始点取为

$$y^{(1)} = x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

在 $y^{(1)}$ 处的函数值 $f(y^{(1)})=17$.

先从 $y^{(1)}$ 出发，沿 $d^{(1)}$ 探测：

$$y^{(1)} + \delta_1 d^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$f(y^{(1)}+\delta_1 d^{(1)})=12<f(y^{(1)})$，探测成功，令

$$y^{(2)} = y^{(1)} + \delta_1 d^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

再从 $y^{(2)}$ 出发，沿 $d^{(2)}$ 探测：

$$y^{(2)} + \delta_2 d^{(2)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$f(y^{(2)}+\delta_2 d^{(2)})=22>f(y^{(2)})$，探测失败，因此取 $y^{(3)}=y^{(2)}$.

经第 1 轮探测，得到 $y^{(3)}=(1,0)^T$. 由于 $f(y^{(3)})=12<f(y^{(1)})$，因此第 1 个探测阶段不结束，需沿原来的方向进行下一轮探测.

继续作下去，经 4 轮探测，完成第 1 个探测阶段，得到 $x^{(2)}=(4,-2)^T$，$x^{(2)}$将作为下一个探测阶段的初点.

第 1 个探测阶段完成后，应定义一组新的单位正交方向，即进行转轴，然后才能开始下一个阶段的探测. 为此先求出沿各坐标方向的步长. 由于

$$x^{(2)} - x^{(1)} = \begin{bmatrix} 4 \\ -2 \end{bmatrix},$$

因此这一探测阶段沿 $d^{(1)}$ 及 $d^{(2)}$ 方向的步长依次为 $\lambda_1=4, \lambda_2=-2$.

根据公式(2.56),令

$$p^{(1)}=\lambda_1 d^{(1)}+\lambda_2 d^{(2)}=\begin{bmatrix}4\\-2\end{bmatrix};$$

$$p^{(2)}=\lambda_2 d^{(2)}=\begin{bmatrix}0\\-2\end{bmatrix}.$$

把 $p^{(1)}$,$p^{(2)}$正交化,令

$$q^{(1)}=p^{(1)}=\begin{bmatrix}4\\-2\end{bmatrix},$$

以及

$$q^{(2)}=p^{(2)}-\frac{q^{(1)T}p^{(2)}}{q^{(1)T}q^{(1)}}q^{(1)}=\begin{bmatrix}-\dfrac{4}{5}\\-\dfrac{8}{5}\end{bmatrix}.$$

把 $q^{(1)}$,$q^{(2)}$单位化,则

$$\bar{d}^{(1)}=\begin{bmatrix}\dfrac{2}{\sqrt{5}}\\-\dfrac{1}{\sqrt{5}}\end{bmatrix},\ \bar{d}^{(2)}=\begin{bmatrix}-\dfrac{1}{\sqrt{5}}\\-\dfrac{2}{\sqrt{5}}\end{bmatrix}.$$

容易验证,这是一组单位正交方向.接下去,从 $y^{(1)}=x^{(2)}$出发,沿新构造的方向进行探测移动,方法与上次迭代相同.继续作下去,所得到的点 $x^{(k)}$会更加接近问题的最优解 $\bar{x}=(3,-2)^T$.

### 2.2.8 可变多面体搜索法

这种方法是由 Nelder 和 Mead 在单纯形方法的基础上修改而成的.他们用 $E^n$ 中有$(n+1)$个顶点的**可变多面体**(variable polyhedron)把具有 $n$ 个独立变量的函数 $f(x)$极小化.其方法是:

先求出单纯形(有 $n+1$ 个顶点的多面体)$n+1$ 个顶点上的函数值,确定出有最大函数值的点(称为最高点)和最小函数值的点(称为最低点).然后通过反射、扩展、压缩方法求出一个较好点,用

它取代最高点，构成新的单纯形，或者通过向最低点收缩形成新的单纯形，用这种方法逼近极小点.

计算步骤：

(1) 给定初始单纯形，其顶点为

$$x^{(i)} \in E^n, i = 1,2,\cdots,n+1,$$

反射系数 $\alpha>0$，扩展系数 $\gamma>1$，压缩系数 $\beta\in(0,1)$，允许误差 $\varepsilon>0$. 计算函数值

$$f(x^{(i)}), i = 1,2,\cdots,n+1,$$

置 $k=1$.

(2) 确定最高点 $x^{(h)}$，次高点 $x^{(g)}$，最低点 $x^{(l)}$，$h,g,l\in\{1,2,\cdots,n+1\}$，使得

$$f(x^{(h)}) = \max\{f(x^{(1)}), f(x^{(2)}),\cdots,f(x^{(n+1)})\},$$

$$f(x^{(g)}) = \max\{f(x^{(i)}) \mid x^{(i)} \neq x^{(h)}\},$$

$$f(x^{(l)}) = \min\{f(x^{(1)}), f(x^{(2)}),\cdots,f(x^{(n+1)})\}.$$

计算除 $x^{(h)}$ 外的 $n$ 个点的形心 $\bar{x}$，令

$$\bar{x} = \frac{1}{n}\left[\sum_{i=1}^{n+1} x^{(i)} - x^{(h)}\right],$$

计算出 $f(\bar{x})$.

(3) 进行反射，令

$$x^{(n+2)} = \bar{x} + \alpha(\bar{x} - x^{(h)}), \qquad (2.59)$$

计算 $f(x^{(n+2)})$.

(4) 若 $f(x^{(n+2)})<f(x^{(l)})$，则进行扩展，令

$$x^{(n+3)} = \bar{x} + \gamma(x^{(n+2)} - \bar{x}), \qquad (2.60)$$

计算 $f(x^{(n+3)})$，转(5)；

若 $f(x^{(l)})\leqslant f(x^{(n+2)})\leqslant f(x^{(g)})$，则置 $x^{(h)}=x^{(n+2)}$，$f(x^{(h)})=f(x^{(n+2)})$，转(7)；

若 $f(x^{(n+2)})>f(x^{(g)})$，则进行压缩，令

$f(x^{(h')}) = \min\{f(x^{(h)}), f(x^{(n+2)})\}$，其中 $h' \in \{h, n+2\}$. 令

$$x^{(n+4)} = \bar{x} + \beta(x^{(h')} - \bar{x}), \tag{2.61}$$

计算 $f(x^{(n+4)})$，转(6).

(5) 若 $f(x^{(n+3)}) < f(x^{(n+2)})$，则置

$$x^{(h)} = x^{(n+3)}, f(x^{(h)}) = f(x^{(n+3)}),$$

转(7)；否则，置 $x^{(h)} = x^{(n+2)}$，$f(x^{(h)}) = f(x^{(n+2)})$，转(7).

(6) 若 $f(x^{(n+4)}) \leqslant f(x^{(h')})$，则置

$$x^{(h)} = x^{(n+4)}, f(x^{(h)}) = f(x^{(n+4)}),$$

进行(7)；否则，进行收缩，令

$$x^{(i)} := x^{(l)} + \frac{1}{2}(x^{(i)} - x^{(l)}), i = 1,2,\cdots,n+1.$$

计算 $f(x^{(i)})$，$i=1,2,\cdots,n+1$，进行(7).

(7) 检验是否满足收敛准则. 若

$$\left\{\frac{1}{n+1}\sum_{i=1}^{n+1}[f(x^{(i)}) - f(\bar{x})]^2\right\}^{\frac{1}{2}} < \varepsilon, \tag{2.62}$$

则停止计算，现行最好点可作为极小点的近似；否则，置 $k := k+1$，返回(2).

**例 2.2.21** 用可变多面体搜索法解下列问题

$$\min \quad f(x) \triangleq (x_1 - 3)^2 + 2(x_2 + 2)^2,$$

取初始单纯形的顶点为

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, x^{(2)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x^{(3)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\alpha = 1, \gamma = 2, \beta = 1/2, \varepsilon = 2.$$

**解** 第 1 次迭代：各顶点处的函数值分别为 $f(x^{(1)}) = 17$，$f(x^{(2)}) = 12$，$f(x^{(3)}) = 27$，记顶点 $x^{(h)} = x^{(3)}$，$x^{(g)} = x^{(1)}$，$x^{(l)} = x^{(2)}$. 将 $x^{(3)}$ 经 $x^{(1)}$ 和 $x^{(2)}$ 的形心进行反射，令

$$\bar{x} = \frac{1}{2}(x^{(1)} + x^{(2)}) = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix},$$

在 $\bar{x}$ 处，$f(\bar{x})=57/4$. 令反射得到的点

$$x^{(4)}=\bar{x}+\alpha(\bar{x}-x^{(3)})=\begin{bmatrix}1\\-1\end{bmatrix},$$

$f(x^{(4)})=6<f(x^{(l)})=12$. 进行扩展，令

$$x^{(5)}=\bar{x}+\gamma(x^{(4)}-\bar{x})=\begin{bmatrix}\frac{3}{2}\\-2\end{bmatrix},$$

由于 $f(x^{(5)})=\frac{9}{4}<f(x^{(4)})$，因此用 $x^{(5)}$替换 $x^{(3)}$，得到新的单纯形. 把 $x^{(5)}$仍记作 $x^{(3)}$，则新的单纯形的顶点是

$$x^{(1)}=\begin{bmatrix}0\\0\end{bmatrix},x^{(2)}=\begin{bmatrix}1\\0\end{bmatrix},x^{(3)}=\begin{bmatrix}\frac{3}{2}\\-2\end{bmatrix}.$$

由于

$$\left\{\frac{1}{3}\sum_{i=1}^{3}[f(x^{(i)})-f(\bar{x})]^2\right\}^{\frac{1}{2}}=7.23>\varepsilon,$$

不满足精度要求，需继续计算.

容易验证，经过 3 次迭代，得到的单纯形其顶点是

$$x^{(1)}=\begin{bmatrix}\frac{5}{2}\\-2\end{bmatrix},x^{(2)}=\begin{bmatrix}\frac{5}{2}\\-3\end{bmatrix},x^{(3)}=\begin{bmatrix}\frac{3}{2}\\-2\end{bmatrix}.$$

各点的函数值为

$$f(x^{(1)})=1/4,f(x^{(2)})=9/4,f(x^{(3)})=9/4.$$

由于

$$\left\{\frac{1}{3}\sum_{i=1}^{3}[f(x^{(i)})-f(\bar{x})]^2\right\}^{1/2}=1.11<\varepsilon,$$

已满足精度要求，得近似解 $x^{(1)}=\left(\frac{5}{2},-2\right)^{\mathrm{T}}$. 实际上，问题的极小点 $\bar{x}=(3,-2)^{\mathrm{T}}$.

### 2.2.9 Powell 法

#### 1 Powell 基本算法

Powell 法是一种有效的直接搜索法，这种方法本质上是共轭方向法.

Powell 方法把整个计算过程分成若干个阶段，每一阶段(一轮迭代)由 $n+1$ 次一维搜索组成. 在算法的每一阶段中，先依次沿着 $n$ 个已知的方向搜索，得一个最好点，然后沿本阶段的初始点与该最好点连线方向进行搜索，求得这一阶段的最好点. 再用最后的搜索方向取代前 $n$ 个方向之一，开始下一阶段的迭代.

计算步骤：

(1) 给定初始点 $\boldsymbol{x}^{(0)}$，$n$ 个线性无关的方向 $\boldsymbol{d}^{(1,1)},\boldsymbol{d}^{(1,2)},\cdots,\boldsymbol{d}^{(1,n)}$，允许误差 $\varepsilon>0$，置 $k=1$.

(2) 置 $\boldsymbol{x}^{(k,0)}=\boldsymbol{x}^{(k-1)}$，从 $\boldsymbol{x}^{(k,0)}$ 出发，依次沿方向 $\boldsymbol{d}^{(k,1)},\boldsymbol{d}^{(k,2)},\cdots,\boldsymbol{d}^{(k,n)}$ 进行搜索，得到点 $\boldsymbol{x}^{(k,1)},\boldsymbol{x}^{(k,2)},\cdots,\boldsymbol{x}^{(k,n)}$，再从 $\boldsymbol{x}^{(k,n)}$ 出发，沿着方向 $\boldsymbol{d}^{(k,n+1)}=\boldsymbol{x}^{(k,n)}-\boldsymbol{x}^{(k,0)}$ 作一维搜索，得到点 $\boldsymbol{x}^{(k)}$.

(3) 若 $\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^{(k-1)}\|<\varepsilon$，则停止计算，得点 $\boldsymbol{x}^{(k)}$；否则，令

$$\boldsymbol{d}^{(k+1,j)}=\boldsymbol{d}^{(k,j+1)},\ j=1,\cdots,n.$$

置 $k:=k+1$，返回(2).

Powell 方法用于极小化正定二次函数时，如果每轮迭代中前 $n$ 个方向线性无关，那么完成 $n$ 个阶段的迭代之后，必能得到 $n$ 个 $\boldsymbol{A}$ 共轭的方向，这里 $\boldsymbol{A}$ 是二次函数的 Hesse 矩阵. 因此，Powell 方法在上述条件下具有二次终止性.

值得注意，在迭代中可能出现这样情形：在某轮迭代中，$n$ 个搜索方向线性相关，由此导致即使对正定二次函数经 $n$ 轮迭代也达不到极小点，甚至任意迭代下去，永远达不到极小点.

**例 2.2.22** 考虑问题

$$\min\ (x_1-x_2+x_3)^2+(-x_1+x_2+x_3)^2+(x_1+x_2-x_3)^2,$$

取初始 $\boldsymbol{x}^{(0)}$ 及搜索方向 $\boldsymbol{d}^{(1,1)},\boldsymbol{d}^{(1,2)},\boldsymbol{d}^{(1,3)}$ 为

$$\boldsymbol{x}^{(0)}=\begin{bmatrix}\frac{1}{2}\\1\\\frac{1}{2}\end{bmatrix},\boldsymbol{d}^{(1,1)}=\begin{bmatrix}1\\0\\0\end{bmatrix},\boldsymbol{d}^{(1,2)}=\begin{bmatrix}0\\1\\0\end{bmatrix},\boldsymbol{d}^{(1,3)}=\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

现在用 Powell 方法求解. 首先置初点 $\boldsymbol{x}^{(1,0)}=\boldsymbol{x}^{(0)}$，从 $\boldsymbol{x}^{(1,0)}$ 出发沿 $\boldsymbol{d}^{(1,1)}$ 搜索，得到点 $\boldsymbol{x}^{(1,1)}=(1/2,1,1/2)^{\mathrm{T}}$.

再从 $\boldsymbol{x}^{(1,1)}$ 出发，沿方向 $\boldsymbol{d}^{(1,2)}$ 搜索，得到点 $\boldsymbol{x}^{(1,2)}=(1/2,1/3,1/2)^{\mathrm{T}}$.

第 3 次搜索，从 $\boldsymbol{x}^{(1,2)}$ 出发，沿着方向 $\boldsymbol{d}^{(1,3)}$ 进行. 得到点 $\boldsymbol{x}^{(1,3)}=(1/2,1/3,5/18)^{\mathrm{T}}$.

这一轮最后一步，从 $\boldsymbol{x}^{(1,3)}$ 出发，沿着方向 $\boldsymbol{d}^{(1,4)}$ 搜索，这个方向定义为

$$\boldsymbol{d}^{(1,4)}=\boldsymbol{x}^{(1,3)}-\boldsymbol{x}^{(1,0)}=\begin{bmatrix}0\\-\frac{2}{3}\\-\frac{2}{9}\end{bmatrix},$$

经搜索得到点 $\boldsymbol{x}^{(1)}=(1/2,1/4,1/4)^{\mathrm{T}}$.

由第 1 轮搜索结果易知，进行第 2 轮搜索时，前 3 个方向 $\boldsymbol{d}^{(1,2)},\boldsymbol{d}^{(1,3)}$ 和 $\boldsymbol{d}^{(1,4)}$ 线性相关. 它们的第 1 个分量都是零，因此 $\boldsymbol{x}^{(1)}$ 出发依次沿这三个方向进行搜索时，第 1 个分量恒为 1/2. 这表明，继续搜索下去永远达不到极小点 $\bar{\boldsymbol{x}}=(0,0,0)^{\mathrm{T}}$.

由上例可见，在 Powell 方法中，保持 $n$ 个方向线性无关十分重要. 然而，Powell 本人已经注意到，即使不像例 2.2.22 那样极端情况，Powell 方法也可能选取接近线性相关的搜索方向，特别是变量很多时更如此. 这种可能性会给收敛性带来严重后果. 为了避免这个困难，人们给出了改进的 Powell 方法.

## 2 改进的 Powell 方法(Sargent 形式)

改进的 Powell 方法与原来方法的主要区别在于替换方向的规则不同. 改进的 Powell 方法,当初始搜索方向线性无关时,能够保证以后每轮迭代中前 $n$ 个方向总是线性无关的. 而且随着迭代的延续,搜索方向接近共轭的程度逐渐增加.

计算步骤:

(1) 给定初始点 $\boldsymbol{x}^{(0)}$,线性无关的方向 $\boldsymbol{d}^{(1,1)},\boldsymbol{d}^{(1,2)},\cdots,\boldsymbol{d}^{(1,n)}$,给定允许误差 $\epsilon>0$,置 $k=1$.

(2) 置 $\boldsymbol{x}^{(k,0)}=\boldsymbol{x}^{(k-1)}$,从 $\boldsymbol{x}^{(k,0)}$ 出发,依次沿方向 $\boldsymbol{d}^{(k,1)},\boldsymbol{d}^{(k,2)},\cdots,\boldsymbol{d}^{(k,n)}$ 作一维搜索,得到点 $\boldsymbol{x}^{(k,1)},\boldsymbol{x}^{(k,2)},\cdots,\boldsymbol{x}^{(k,n)}$. 求指标 $m$,使得

$$f(\boldsymbol{x}^{(k,m-1)})-f(\boldsymbol{x}^{(k,m)})=\max_{j=1,2,\cdots,n}\{f(\boldsymbol{x}^{(k,j-1)})-f(\boldsymbol{x}^{(k,j)})\}.$$

令 $\boldsymbol{d}^{(k,n+1)}=\boldsymbol{x}^{(k,n)}-\boldsymbol{x}^{(k,0)}$,若 $\|\boldsymbol{x}^{(k,n)}-\boldsymbol{x}^{(k,0)}\|\leqslant\epsilon$,则停止计算;否则,进行(3).

(3) 求 $\lambda_{n+1}$,使得

$$f(\boldsymbol{x}^{(k,0)}+\lambda_{n+1}\boldsymbol{d}^{(k,n+1)})=\min_{\lambda}f(\boldsymbol{x}^{(k,0)}+\lambda\boldsymbol{d}^{(k,n+1)}).$$

令 $\boldsymbol{x}^{(k+1,0)}=\boldsymbol{x}^{(k)}=\boldsymbol{x}^{(k,0)}+\lambda_{n+1}\boldsymbol{d}^{(k,n+1)}$,若 $\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^{(k-1)}\|\leqslant\epsilon$,则停止计算,得到点 $\boldsymbol{x}^{(k)}$;否则,进行(4).

(4) 若

$$|\lambda_{n+1}|>\left[\frac{f(\boldsymbol{x}^{(k,0)})-f(\boldsymbol{x}^{(k+1,0)})}{f(\boldsymbol{x}^{(k,m-1)})-f(\boldsymbol{x}^{(k,m)})}\right]^{1/2},\qquad(2.63)$$

则令
$$\boldsymbol{d}^{(k+1,j)}=\boldsymbol{d}^{(k,j)},\ j=1,\cdots,m-1;$$
$$\boldsymbol{d}^{(k+1,j)}=\boldsymbol{d}^{(k,j+1)},\ j=m,\cdots,n.$$

置 $k:=k+1$,转(2);否则,令

$$\boldsymbol{d}^{(k+1,j)}=\boldsymbol{d}^{(k,j)},\ j=1,\cdots,n,$$

置 $k:=k+1$,转(2).

改进的 Powell 方法不再具有二次终止性,但它的计算效果仍然令人满意.

## 2.3 约束非线性规划方法

### 2.3.1 Zoutendijk 可行方向法

可行方向法可看作无约束下降算法的自然推广，其典型策略是从可行点出发，沿着下降的可行方向进行搜索，求出使目标函数值下降的新的可行点. 算法的主要步骤是选择搜索方向和确定沿此方向移动的步长. 搜索方向的选择方式不同就形成各种可行方向法. 下面给出 Zoutendijk 可行方向法.

**1 线性约束情形**

考虑非线性规划问题：

$$\left.\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \\ & \boldsymbol{E}\boldsymbol{x} = \boldsymbol{e}, \end{aligned}\right\} \tag{2.64}$$

其中 $f(\boldsymbol{x})$ 是可微函数，$\boldsymbol{A}$ 为 $m\times n$ 矩阵，$\boldsymbol{E}$ 为 $l\times n$ 矩阵，$\boldsymbol{x}\in E^n$，$\boldsymbol{b}$ 和 $\boldsymbol{e}$ 分别为 $m$ 维及 $l$ 维列向量. 可行域记作

$$S = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b}, \boldsymbol{E}\boldsymbol{x} = \boldsymbol{e}, \boldsymbol{x} \in E^n\}.$$

用可行方向法求解上述问题，整个过程分作两步. 假设 $\boldsymbol{x}^{(k)}\in S$ 已知. 第一步，在 $\boldsymbol{x}^{(k)}$ 求一可行下降方向 $\boldsymbol{d}^{(k)}$；第二步，从 $\boldsymbol{x}^{(k)}$ 出发，沿方向 $\boldsymbol{d}^{(k)}$ 搜索，求出步长 $\lambda_k$，使得后继点 $\boldsymbol{x}^{(k+1)}\in S$.

为求可行下降方向 $\boldsymbol{d}^{(k)}$，需解下列线性规划：

$$\left.\begin{aligned} \min \quad & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d}; \\ \text{s.t.} \quad & \boldsymbol{A}_1\boldsymbol{d} \geqslant \boldsymbol{0}, \\ & \boldsymbol{E}\boldsymbol{d} = \boldsymbol{0}, \\ & |d_j| \leqslant 1, j = 1,\cdots,n. \end{aligned}\right\} \tag{2.65}$$

其中 $\boldsymbol{A}_1$ 是不等式约束 $\boldsymbol{A}\boldsymbol{x}\geqslant\boldsymbol{b}$ 中，在 $\boldsymbol{x}^{(k)}$ 处起作用约束的系数矩阵. 具体地，假设在 $\boldsymbol{x}^{(k)}$ 点可以写作

$$A=\begin{bmatrix}A_1\\A_2\end{bmatrix},\ b=\begin{bmatrix}b_1\\b_2\end{bmatrix},$$

使得 $A_1x^{(k)}=b_1$，$A_2x^{(k)}>b_2$. 约束条件 $|d_j|\leqslant 1$，是为了获得一个有限解.

对于问题(2.65)，由于 $d=0$ 是可行解，因此目标函数的最优值必定小于或等于零. 如果目标函数 $\nabla f(x^{(k)})^{\mathrm{T}}d$ 的最优值小于零，则得到下降可行方向 $d$；如果目标函数 $\nabla f(x^{(k)})^{\mathrm{T}}d$ 的最优值为零，则如下面定理所述，$x$ 是 Kuhn-Tucker 点.

**定理 2.3.1** 考虑(2.64)，设 $x$ 是可行解，在点 $x$ 处有 $A_1x=b_1$，$A_2x>b_2$，其中

$$A=\begin{bmatrix}A_1\\A_2\end{bmatrix},\ b=\begin{bmatrix}b_1\\b_2\end{bmatrix},$$

则 $x$ 为 Kuhn-Tucker 点的充要条件是(2.65)的目标函数最优值为零.

根据定理，求解(2.65)的结果，或者是得到下降可行方向，或者得到 Kuhn-Tucker 点.

方向 $d^{(k)}$ 确定后，再求沿此方向移动的步长 $\lambda_k$. 为此，求解下列问题：

$$\begin{aligned}\min\quad & f(x^{(k)}+\lambda d^{(k)});\\ \text{s.t.}\quad & 0\leqslant\lambda\leqslant\lambda_{\max}.\end{aligned}\tag{2.66}$$

其中 $\lambda_{\max}$ 是步长上限，加此上限的目的是使后继点为可行点，即使得

$$x^{(k)}+\lambda_k d^{(k)}\in S.$$

确定 $\lambda_{\max}$ 的方法如下：

$$\hat{b}=b_2-A_2x^{(k)}.\tag{2.67}$$

$$\hat{d}=A_2d^{(k)}.\tag{2.68}$$

$$\lambda_{\max}=\begin{cases}\min\left\{\dfrac{\hat{b}_i}{\hat{d}_i}\middle|\hat{d}_i<0\right\}, & \text{当 } \hat{\boldsymbol{d}} \ngeq 0;\\ \infty, & \text{当 } \hat{\boldsymbol{d}} \geqslant 0.\end{cases} \tag{2.69}$$

其中 $\hat{b}_i$，$\hat{d}_i$ 分别为 $\hat{\boldsymbol{b}}$ 和 $\hat{\boldsymbol{d}}$ 的第 $i$ 个分量.

计算步骤：

(1) 给定初始可行点 $\boldsymbol{x}^{(1)}$，置 $k=1$.

(2) 在点 $\boldsymbol{x}^{(k)}$ 处，把 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 分解成 $\begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{bmatrix}$ 和 $\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{bmatrix}$，使得 $\boldsymbol{A}_1\boldsymbol{x}^{(k)}=\boldsymbol{b}_1$，$\boldsymbol{A}_2\boldsymbol{x}^{(k)}>\boldsymbol{b}_2$. 计算 $\nabla f(\boldsymbol{x}^{(k)})$.

(3) 求解线性规划问题

$$\begin{aligned}\min\quad & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d};\\ \text{s.t.}\quad & \boldsymbol{A}_1\boldsymbol{d}\geqslant \boldsymbol{0},\\ & \boldsymbol{E}\boldsymbol{d}=\boldsymbol{0},\\ & -1\leqslant d_j\leqslant 1,\ j=1,\cdots,n.\end{aligned}$$

得到最优解 $\boldsymbol{d}^{(k)}$.

(4) 如果 $\nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d}^{(k)}=0$，则停止计算，$\boldsymbol{x}^{(k)}$ 为 Kuhn-Tucker 点；否则，进行(5).

(5) 利用(2.67)至(2.69)计算 $\lambda_{\max}$，然后在 $[0,\lambda_{\max}]$ 上作一维搜索：

$$\begin{aligned}\min\quad & f(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)});\\ \text{s.t.}\quad & 0\leqslant\lambda\leqslant\lambda_{\max}.\end{aligned}$$

得到最优解 $\lambda_k$，令

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k\boldsymbol{d}^{(k)}.$$

(6) 置 $k:=k+1$，返回(2).

**例 2.3.2** 用 Zoutendijk 可行方向法解下列问题

$$\begin{aligned}
\min \quad & x_1^2 + x_2^2 - 2x_1 - 4x_2 + 6; \\
\text{s.t.} \quad & -2x_1 + x_2 + 1 \geqslant 0, \\
& -x_1 - x_2 + 2 \geqslant 0, \\
& x_1 \geqslant 0, \\
& x_2 \geqslant 0.
\end{aligned}$$

取初始可行点 $\boldsymbol{x}^{(1)}=(0,0)^{\mathrm{T}}$.

**解** 第1次迭代:

$\boldsymbol{x}^{(1)}$处目标函数的梯度$\nabla f(\boldsymbol{x}^{(1)})=(-2,-4)^{\mathrm{T}}$,起作用约束和不起作用约束的系数矩阵及右端分别为

$$\boldsymbol{A}_1=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix},\quad \boldsymbol{A}_2=\begin{bmatrix}-2 & 1\\-1 & -1\end{bmatrix},$$

$$\boldsymbol{b}_1=\begin{bmatrix}0\\0\end{bmatrix},\quad \boldsymbol{b}_2=\begin{bmatrix}-1\\-2\end{bmatrix}.$$

求在$\boldsymbol{x}^{(1)}$处的下降可行方向,为此需解线性规划问题:

$$\begin{aligned}
\min \quad & \nabla f(\boldsymbol{x}^{(1)})^{\mathrm{T}}\boldsymbol{d}; \\
\text{s.t.} \quad & \boldsymbol{A}_1\boldsymbol{d} \geqslant 0, \\
& |d_j| \leqslant 1, j=1,2.
\end{aligned}$$

即求解

$$\begin{aligned}
\min \quad & -2d_1 - 4d_2; \\
\text{s.t.} \quad & d_1 \geqslant 0, \\
& d_2 \geqslant 0, \\
& -1 \leqslant d_1 \leqslant 1, \\
& -1 \leqslant d_2 \leqslant 1.
\end{aligned}$$

用单纯形方法解此线性规划,得到最优解

$$\boldsymbol{d}^{(1)}=(1,1)^{\mathrm{T}}.$$

再求步长 $\lambda_1$:

$$\hat{\boldsymbol{d}} = \boldsymbol{A}_2\boldsymbol{d}^{(1)} = (-1, -2)^{\mathrm{T}},$$

$$\hat{\boldsymbol{b}} = \boldsymbol{b}_2 - \boldsymbol{A}_2\boldsymbol{x}^{(1)} = (-1, -2)^{\mathrm{T}},$$

$$\lambda_{\max} = \min\left\{\frac{-1}{-1}, \frac{-2}{-2}\right\} = 1.$$

解一维搜索问题：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}^{(1)} + \lambda\boldsymbol{d}^{(1)}) \triangleq 2\lambda^2 - 6\lambda + 6; \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant 1. \end{aligned}$$

解得 $\lambda_1 = 1$. 经第 1 次迭代得到点

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \lambda_1\boldsymbol{d}^{(1)} = (1, 1)^{\mathrm{T}}.$$

后面迭代，按同样方式进行.

第 2 次迭代中，求得可行方向 $\boldsymbol{d}^{(2)} = (-1, 1)^{\mathrm{T}}$. 从 $\boldsymbol{x}^{(2)}$ 出发，沿 $\boldsymbol{d}^{(2)}$ 方向进行一维搜索，得到步长 $\lambda_2 = 1/2$. 因此

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \lambda_2\boldsymbol{d}^{(2)} = (1/2, 3/2)^{\mathrm{T}}.$$

第 3 次迭代，从 $\boldsymbol{x}^{(3)}$ 出发. 在求可行方向时得到 $\boldsymbol{d}^{(3)} = (0, 0)^{\mathrm{T}}$.

根据定理 2.3.1，$\boldsymbol{x}^{(3)} = (1/2, 3/2)^{\mathrm{T}}$ 一定是 Kuhn-Tucker 点. 由于例是凸规划，因此 $\boldsymbol{x}^{(3)}$ 是最优解，目标函数的最优值 $f_{\min} = f(\boldsymbol{x}^{(3)}) = 3/2$.

### 2 非线性约束情形

考虑问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \geqslant 0, i = 1, \cdots, m. \end{aligned} \tag{2.70}$$

其中 $\boldsymbol{x} \in E^n$，$f(\boldsymbol{x})$，$g_i(\boldsymbol{x})$ 均为可微函数.

为了确定在点 $\boldsymbol{x}$ 处的可行方向，需要求解下列线性规划问题：

$$\begin{aligned} \min \quad & z; \\ \text{s.t.} \quad & \nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0, \\ & \nabla g_i(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} + z \geqslant 0, i \in I, \\ & |d_j| \leqslant 1, j = 1, \cdots, n. \end{aligned} \tag{2.71}$$

其中 $I = \{i \mid g_i(\boldsymbol{x}) = 0\}$

设(2.71)的最优解为$(\bar{z},\bar{\boldsymbol{d}})$. 如果$\bar{z}<0$,则$\bar{\boldsymbol{d}}$是在$\boldsymbol{x}$处的下降可行方向;如果$\bar{z}=0$,则$\boldsymbol{x}$为Fritz John点.

运用可行方向法求$\boldsymbol{x}^{(k)}$的后继点时,先解线性规划(2.71)(其中$\boldsymbol{x}$以$\boldsymbol{x}^{(k)}$代入),求出在点$\boldsymbol{x}^{(k)}$的可行下降方向$\boldsymbol{d}^{(k)}$,然后确定沿$\boldsymbol{d}^{(k)}$方向移动的步长$\lambda_k$. 为此需要求解下列一维搜索问题:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{d}^{(k)}); \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant \lambda_{\max}. \end{aligned} \tag{2.72}$$

其中

$$\lambda_{\max} = \sup\{\lambda \mid g_i(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{d}^{(k)}) \geqslant 0, i=1,\cdots,m\}. \tag{2.73}$$

计算步骤:

(1) 给定初始可行点$\boldsymbol{x}^{(1)}$,置$k=1$.

(2) 令$I=\{i \mid g_i(\boldsymbol{x}^{(k)})=0\}$,解线性规划问题:

$$\begin{aligned} \min \quad & z; \\ \text{s.t.} \quad & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0, \\ & \nabla g_i(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d} + z \geqslant 0, i \in I, \\ & -1 \leqslant d_j \leqslant 1, j=1,\cdots,n. \end{aligned}$$

得到最优解$(z_k,\boldsymbol{d}^{(k)})$.

若$z_k=0$,则停止计算,$\boldsymbol{x}^{(k)}$为Fritz John点;否则,进行(3).

(3) 求解一维搜索问题:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{d}^{(k)}); \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant \lambda_{\max}, \end{aligned}$$

其中$\lambda_{\max}$由(2.73)确定,得到最优解$\lambda_k$.

(4) 令$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k \boldsymbol{d}^{(k)}$,置$k:=k+1$,返回(2).

### 3 Topkis-Veinott 修正

Zoutendijk 法产生的序列可能不收敛于 Kuhn-Tucker 点,因此 Topkis 和 Veinott 对这种方法做了改进,把求方向的线性规划改写成:

$$
\begin{aligned}
\min \quad & z; \\
\text{s.t.} \quad & \nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0, \\
& \nabla g_i(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d} + z \geqslant -g_i(\boldsymbol{x}), i = 1,\cdots,m, \\
& -1 \leqslant d_j \leqslant 1, j = 1,\cdots,n.
\end{aligned}
$$

经过修改，紧约束和非紧约束在确定下降可行方向中均起作用，并且在接近非紧约束边界时，不至发生方向突然改变. Topkis-Veinott 算法产生的序列$\{\boldsymbol{x}^{(k)}\}$，其任一聚点是 Fritz John 点. 即对任一聚点$\bar{\boldsymbol{x}}$，存在不全为零的非负数$w_0$，及$w_i(i\in \boldsymbol{I})$，使得

$$
w_0\nabla f(\bar{\boldsymbol{x}}) - \sum_{i\in \boldsymbol{I}} w_i \nabla g_i(\bar{\boldsymbol{x}}) = 0.
$$

计算步骤：

(1) 给定初始可行点$\boldsymbol{x}^{(1)}$，置$k=1$.

(2) 求解线性规划

$$
\begin{aligned}
\max \quad & z; \\
\text{s.t.} \quad & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0, \\
& \nabla g_i(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{d} + z \geqslant -g_i(\boldsymbol{x}^{(k)}), i = 1,\cdots,m, \\
& -1 \leqslant d_j \leqslant 1, j = 1,\cdots,n.
\end{aligned}
$$

得到最优解$(z_k,\boldsymbol{d}^{(k)})$.

(3) 若$z_k=0$，则停止计算，这时，$\boldsymbol{x}^{(k)}$为 Fritz John 点；否则，进行(4).

(4) 求步长$\lambda_k$：

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}^{(k)} + \lambda \boldsymbol{d}^{(k)}); \\
\text{s.t.} \quad & 0 \leqslant \lambda \leqslant \lambda_{\max},
\end{aligned}
$$

其中$\lambda_{\max}=\sup\{\lambda \mid g_i(\boldsymbol{x}^{(k)}+\lambda \boldsymbol{d}^{(k)})\geqslant 0, i=1,\cdots,m\}$，求得最优解$\lambda_k$.

(5) 令$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k \boldsymbol{d}^{(k)}$，置$k:=k+1$，返回(2).

### 2.3.2 Rosen 梯度投影法

考虑问题

$$\begin{aligned}\min\quad & f(\boldsymbol{x});\\ \text{s.t.}\quad & \boldsymbol{Ax} \geqslant \boldsymbol{b},\\ & \boldsymbol{Ex} = \boldsymbol{e}.\end{aligned} \tag{2.74}$$

其中 $f(\boldsymbol{x})$是可微函数,$\boldsymbol{A}$ 为 $m\times n$ 矩阵,$\boldsymbol{E}$ 为 $l\times n$ 矩阵.

为了介绍 Rosen **梯度投影法**(gradient projection method),先给出**投影矩阵**(projection matrix)概念.

**定义 2.3.3** 设 $\boldsymbol{P}$ 为 $n$ 阶矩阵,若 $\boldsymbol{P}=\boldsymbol{P}^{\mathrm{T}}$ 且 $\boldsymbol{P}^2=\boldsymbol{P}$,则称 $\boldsymbol{P}$ 为**投影矩阵**.

**例 2.3.4** 设 $\boldsymbol{M}$ 为 $m\times n$ 矩阵,秩为 $m$,令

$$\boldsymbol{Q} = \boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MM}^{\mathrm{T}})^{-1}\boldsymbol{M}, \tag{2.75}$$

及

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MM}^{\mathrm{T}})^{-1}\boldsymbol{M}, \tag{2.76}$$

则 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 均为投影矩阵.

投影矩阵具有下列性质:

(1) 若 $\boldsymbol{P}$ 为投影矩阵,则 $\boldsymbol{P}$ 为半正定矩阵.

(2) $\boldsymbol{P}$ 为投影矩阵的充要条件是 $\boldsymbol{Q}=\boldsymbol{I}-\boldsymbol{P}$ 为投影矩阵.

(3) 设 $\boldsymbol{P}$ 和 $\boldsymbol{Q}=\boldsymbol{I}-\boldsymbol{P}$ 是 $n$ 阶投影矩阵,则

$$L = \{\boldsymbol{Px} \mid \boldsymbol{x} \in E^n\}$$

与

$$L^{\perp} = \{\boldsymbol{Qx} \mid \boldsymbol{x} \in E^n\}$$

是正交线性子空间,且任一 $\boldsymbol{x}\in E^n$ 可唯一分解成 $\boldsymbol{x}=\boldsymbol{p}+\boldsymbol{q}$,$\boldsymbol{p}\in L$,$\boldsymbol{q}\in L^{\perp}$.

梯度投影法的基本思想仍然是从可行点出发,沿可行方向进行搜索.当迭代出发点在可行域内部时,沿负梯度方向搜索.当迭代出发点在某些约束的边界上时,将该点处的负梯度投影到 $\boldsymbol{M}$ 的**零空间**(null space)(即 $\boldsymbol{Mx}=\boldsymbol{0}$ 的解空间).$\boldsymbol{M}$ 是以起作用约束的梯度为行构成的矩阵.下面定理 2.3.5 表明,这样的投影是下降可行方向.再沿此方向进行搜索.因此,Rosen 梯度投影法也是一种

可行方向法.

**定理 2.3.5** 设 $\boldsymbol{x}$ 是问题(2.74)的可行解，在 $\boldsymbol{x}$ 处有 $\boldsymbol{A}_1\boldsymbol{x}=\boldsymbol{b}_1$，$\boldsymbol{A}_2\boldsymbol{x}>\boldsymbol{b}_2$，其中

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{bmatrix},\quad \boldsymbol{b}=\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{bmatrix},$$

$$\boldsymbol{M}\triangleq\begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{E}\end{bmatrix}\quad \text{为满秩矩阵},$$

$\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}})^{-1}\boldsymbol{M}$，

$$\boldsymbol{P}\nabla f(\boldsymbol{x})\neq\boldsymbol{0},\ \boldsymbol{d}\triangleq-\boldsymbol{P}\nabla f(\boldsymbol{x}),$$

则 $\boldsymbol{d}$ 是下降可行方向.

这个定理，在 $\boldsymbol{P}\nabla f(\boldsymbol{x})\neq\boldsymbol{0}$ 的假设下，给出用投影求下降可行方向的一种方法.

当 $\boldsymbol{P}\nabla f(\boldsymbol{x})=\boldsymbol{0}$ 时，有两种可能，或者 $\boldsymbol{x}$ 是 Kuhn-Tucker 点，或者可以构造新的投影矩阵，以便求得下降可行方向.

**定理 2.3.6** 设 $\boldsymbol{x}$ 是(2.74)的一个可行解，在点 $\boldsymbol{x}$ 处，有 $\boldsymbol{A}_1\boldsymbol{x}=\boldsymbol{b}_1$，$\boldsymbol{A}_2\boldsymbol{x}>\boldsymbol{b}_2$，其中

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{A}_2\end{bmatrix},\ \boldsymbol{b}=\begin{bmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{bmatrix}.$$

$\boldsymbol{M}\triangleq\begin{bmatrix}\boldsymbol{A}_1\\ \boldsymbol{E}\end{bmatrix}$为满秩矩阵，$\boldsymbol{P}\triangleq\boldsymbol{I}-\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}})^{-1}\boldsymbol{M}$，

$$\boldsymbol{w}=(\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}})^{-1}\boldsymbol{M}\nabla f(\boldsymbol{x})=\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{v}\end{bmatrix},$$

其中 $\boldsymbol{u}$ 和 $\boldsymbol{v}$ 分别对应于 $\boldsymbol{A}_1$ 和 $\boldsymbol{E}$，设 $\boldsymbol{P}\nabla f(\boldsymbol{x})=0$，则

(1) 如果 $\boldsymbol{u}\geqslant 0$，那么 $\boldsymbol{x}$ 为 Kuhn-Tucker 点；

(2) 如果 $\boldsymbol{u}$ 中含有负分量，不妨假设 $u_j<0$，这时从 $\boldsymbol{A}_1$ 中去掉 $u_j$ 对应的行，得到 $\hat{\boldsymbol{A}}_1$，

$$\hat{M} \triangleq \begin{bmatrix} \hat{A}_1 \\ E \end{bmatrix},$$

$$\hat{P} \triangleq I - \hat{M}^{\mathrm{T}}(\hat{M}\hat{M}^{\mathrm{T}})^{-1}\hat{M}, \quad d \triangleq -\hat{P}\nabla f(x)$$

那么 $d$ 为下降可行方向.

计算步骤:

(1) 给定初始可行点 $x^{(1)}$,置 $k=1$.

(2) 在点 $x^{(k)}$ 处,将 $A$ 和 $b$ 分解成 $\begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$,使得 $A_1 x^{(k)} = b_1$, $A_2 x^{(k)} > b_2$.

(3) $M \triangleq \begin{bmatrix} A_1 \\ E \end{bmatrix}$,如果 $M$ 是空的,则 $P \triangleq I$(单位矩阵);否则, $P \triangleq I - M^{\mathrm{T}}(MM^{\mathrm{T}})^{-1}M$.

(4) $d^{(k)} \triangleq -P\nabla f(x^{(k)})$. 若 $d^{(k)} \neq 0$ 则转(6);若 $d^{(k)} = 0$,则进行(5).

(5) 若 $M$ 是空的,则停止计算,得到 $x^{(k)}$;否则,

$$w \triangleq (MM^{\mathrm{T}})^{-1}M\nabla f(x^{(k)}) = \begin{bmatrix} u \\ v \end{bmatrix},$$

如果 $u \geqslant 0$,则停止计算,$x^{(k)}$ 为 K-T 点;如果 $u$ 包含负分量,则选择一个负分量,比如 $u_j$,修正 $A_1$,去掉 $A_1$ 中对应 $u_j$ 的行,返回(3).

(6) 求步长 $\lambda_k$,解下列问题:

$$\begin{aligned} \min \quad & f(x^{(k)} + \lambda d^{(k)}); \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant \lambda_{\max}, \end{aligned}$$

其中 $\lambda_{\max}$ 由(2.69)确定. 得解 $\lambda_k$ 后,

$$x^{(k+1)} \triangleq x^{(k)} + \lambda_k d^{(k)},$$

置 $k:=k+1$,返回(2).

**例 2.3.7** 用 Rosen 梯度投影法求解下列问题:

$$\min \quad f(x) \triangleq 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2;$$

$$\begin{aligned} \text{s.t.}\quad & -x_1 - x_2 \geqslant -2, \\ & -x_1 - 5x_2 \geqslant -5, \\ & x_1 \geqslant 0 \\ & x_2 \geqslant 0, \end{aligned}$$

初点取作 $\boldsymbol{x}^{(1)}=(0,0)^{\mathrm{T}}$.

**解** 第1次迭代：在点 $\boldsymbol{x}^{(1)}$ 处梯度 $\nabla f(\boldsymbol{x}^{(1)})=\begin{bmatrix}-4\\-6\end{bmatrix}$.

在 $\boldsymbol{x}^{(1)}$ 处起作用约束指标集 $I=\{3,4\}$，即 $x_1\geqslant0$ 和 $x_2\geqslant0$ 是在 $\boldsymbol{x}^{(1)}$ 处的起作用约束. 其余约束是不起作用约束，因此将约束系数矩阵 $\boldsymbol{A}$ 和右端 $\boldsymbol{b}$ 分解为

$$\boldsymbol{A}_1=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix},\ \boldsymbol{A}_2=\begin{bmatrix}-1 & -1\\-1 & -5\end{bmatrix},$$

$$\boldsymbol{b}_1=\begin{bmatrix}0\\0\end{bmatrix},\quad \boldsymbol{b}_2=\begin{bmatrix}-2\\-5\end{bmatrix}.$$

投影矩阵

$$\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{A}_1^{\mathrm{T}}(\boldsymbol{A}_1\boldsymbol{A}_1^{\mathrm{T}})^{-1}\boldsymbol{A}_1=\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}.$$

负梯度方向在 $\boldsymbol{A}_1$ 的零空间上的投影

$$\boldsymbol{d}^{(1)}=-\boldsymbol{P}\nabla f(\boldsymbol{x}^{(1)})=\begin{bmatrix}0\\0\end{bmatrix}.$$

计算向量 $\boldsymbol{w}$：

$$\boldsymbol{w}=(\boldsymbol{A}_1\boldsymbol{A}_1^{\mathrm{T}})^{-1}\boldsymbol{A}_1\nabla f(\boldsymbol{x}^{(1)})=\begin{bmatrix}u_1\\u_2\end{bmatrix}=\begin{bmatrix}-4\\-6\end{bmatrix},$$

修正 $\boldsymbol{A}_1$，去掉 $\boldsymbol{A}_1$ 中对应 $u_2=-6$ 的行，即第2行，得到 $\hat{\boldsymbol{A}}_1=(1,0)$.

再求投影矩阵 $\hat{\boldsymbol{P}}$：

$$\hat{\boldsymbol{P}}=\boldsymbol{I}-\hat{\boldsymbol{A}}_1^{\mathrm{T}}(\hat{\boldsymbol{A}}_1\hat{\boldsymbol{A}}_1^{\mathrm{T}})^{-1}\hat{\boldsymbol{A}}_1=\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

令投影方向

$$\hat{\boldsymbol{d}}^{(1)} = -\hat{\boldsymbol{P}} \nabla f(\boldsymbol{x}^{(1)}) = \begin{bmatrix} 0 \\ 6 \end{bmatrix},$$

求沿方向 $\hat{\boldsymbol{d}}^{(1)}$ 移动的步长 $\lambda_1$，求解

$$\begin{aligned} \min \quad & f(\boldsymbol{x}^{(1)} + \lambda \hat{\boldsymbol{d}}^{(1)}); \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant \lambda_{\max}. \end{aligned} \tag{2.77}$$

为此先求步长上限 $\lambda_{\max}$. 由于

$$\hat{\boldsymbol{b}} = \boldsymbol{b}_2 - \boldsymbol{A}_2 \boldsymbol{x}^{(1)} = \begin{bmatrix} -2 \\ -5 \end{bmatrix}, \quad \hat{\boldsymbol{d}} = \boldsymbol{A}_2 \hat{\boldsymbol{d}}^{(1)} = \begin{bmatrix} -6 \\ -30 \end{bmatrix},$$

根据(2.69)，有

$$\lambda_{\max} = \min\left\{\frac{-2}{-6}, \frac{-5}{-30}\right\} = \frac{1}{6}$$

这样，(2.77)即

$$\begin{aligned} \min \quad & 72\lambda^2 - 36\lambda; \\ \text{s.t.} \quad & 0 \leqslant \lambda \leqslant 1/6. \end{aligned}$$

解得 $\lambda_1 = 1/6$. 经第 1 次迭代，得到

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \lambda_1 \hat{\boldsymbol{d}}^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

第 2 次迭代与上次类似，经迭代得到

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \lambda_2 \hat{\boldsymbol{d}}^{(2)} = \begin{bmatrix} \frac{35}{31} \\ \frac{24}{31} \end{bmatrix}.$$

第 3 次迭代，以 $\boldsymbol{x}^{(3)}$ 为始点. 经简单计算易知，在 $\boldsymbol{x}^{(3)}$ 处投影方向

$$\boldsymbol{d}^{(3)} = -\boldsymbol{P} \nabla f(\boldsymbol{x}^{(3)}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

而且 $w=(A_1A_1^{T})^{-1}A_1\nabla f(x^{(3)})=\frac{32}{31}>0$.

根据定理(2.3.6)，$x^{(3)}$ 为 Kuhn-Tucker 点．由于本例为凸规划，Kuhn-Tucker 点是整体最优解．

### 2.3.3 既约梯度法

#### 1 Wolfe 既约梯度法

Wolfe 于 1963 年提出产生下降可行方向的另一种方法，称为**既约梯度法**(reduced gradient method)，下面简述算法原理．

考虑具有线性约束的非线性规划问题

$$\begin{aligned} \min \quad & f(x); \\ \text{s.t.} \quad & Ax = b, \\ & x \geqslant 0. \end{aligned} \tag{2.78}$$

其中 $A$ 为 $m\times n$ 矩阵，秩为 $m$，$b$ 是 $m$ 维列向量，$f$ 是 $E^n$ 上的连续可微函数．假设 $A$ 的任意 $m$ 个列均线性无关，并且约束条件的每个基本可行解均有 $m$ 个正分量，在此假设下，每个可行解至少有 $m$ 个正分量，至多有$(n-m)$个零分量．

Wolfe 既约梯度法的基本思想是，把变量区分为**基变量**(basic variable)，有 $m$ 个，和**非基变量**(nonbasic variable)，它们之间的关系由约束条件 $Ax=b$ 确定，将基变量用非基变量表示，并从目标函数中消去基变量，得到以非基变量为自变量的简化的目标函数，进而利用此函数的负梯度构造下降可行方向．简化目标函数关于非基变量的梯度称为目标函数的**既约梯度**(reduced gradient)．算法的中心问题是使用既约梯度构造搜索方向．

先给出既约梯度表达式．

可令

$$A=(B,N),\ x=\begin{bmatrix} x_B \\ x_N \end{bmatrix},$$

其中 $B$ 是 $m\times m$ 阶可逆矩阵，$x_B$ 和 $x_N$ 分别是由基变量和非基变量构成的向量.(2.78)的等式约束中，将 $x_B$ 用 $x_N$ 表示：

$$x_B = B^{-1}b - B^{-1}Nx_N \tag{2.79}$$

$$F(x_N) = f(x_B(x_N), x_N).$$

原来问题简化为仅在变量非负的限制下极小化 $F(x_N)$. 于是，问题(2.78)可以转化为求解下列问题：

$$\left.\begin{array}{ll}\min & F(x);\\ \text{s.t.} & x_B, x_N \geqslant 0.\end{array}\right\} \tag{2.80}$$

利用复合函数求导数法则，可求得 $F(x_N)$的梯度，即 $f(x)$的既约梯度.

$$\begin{aligned}\gamma(x_N) &= \nabla F(x_N)\\ &= \nabla_{x_N} f(x_B(x_N), x_N) - (B^{-1}N)^{\mathrm{T}}\nabla_{x_B} f(x_B(x_N), x_N).\end{aligned} \tag{2.81}$$

下面利用既约梯度构造搜索方向：

令搜索方向

$$d^{(k)} = \begin{bmatrix} d_B^{(k)} \\ d_N^{(k)} \end{bmatrix},$$

其中 $d_B^{(k)}$ 和 $d_N^{(k)}$ 分别对应基变量和非基变量.

为使目标函数值下降，$d_N^{(k)}$ 应取负既约梯度方向. 但是当某个分量 $x_{N_j}=0$ 且 $\gamma_j(x_N)>0$ 时，沿负既约梯度方向移动将破坏可行性，因此定义 $d_N^{(k)}$，使得

$$d_{N_j}^{(k)} = \begin{cases} -x_{N_j}^{(k)}\gamma_j(x_N^{(k)}), & 当\ \gamma_j(x_N^{(k)}) > 0,\\ -\gamma_j(x_N^{(k)}), & 当\ \gamma_j(x_N^{(k)}) \leqslant 0.\end{cases} \tag{2.82}$$

并令

$$d_B^{(k)} = -B^{-1}Nd_N^{(k)}. \tag{2.83}$$

由于 $x_B^{(k)}>0$，因此 $d_B^{(k)}$ 对于 $x_B^{(k)}$ 也一定是可行方向. 最终得到

$$d^{(k)} = \begin{bmatrix} -B^{-1}Nd_N^{(k)} \\ d_N^{(k)} \end{bmatrix}. \tag{2.84}$$

从 $x^{(k)}$ 出发沿方向 $d^{(k)}$ 移动的步长 $\lambda$，其上限由下式确定：

$$\lambda_{\max} = \begin{cases} \infty, & \text{当 } d^{(k)} \geqslant 0, \\ \min\left\{ -\dfrac{x_j^{(k)}}{d_j^{(k)}} \,\middle|\, d_j^{(k)} < 0 \right\}, & \text{其它}. \end{cases} \tag{2.85}$$

容易证明，按照上述方式构造的方向 $d$ 为零向量时，相应的点 $x$ 为 Kuhn-Tucker 点；$d$ 为非零向量时，它必是下降可行方向.

**定理 2.3.8** 设 $x$ 是问题(2.78)的可行解，$A=(B,N)$ 是 $m\times n$ 矩阵，$B$ 为 $m$ 阶可逆矩阵，$x=(x_B^T,x_N^T)^T$，$x_B>0$，函数 $f$ 在点 $x$ 处可微，又设 $d$ 是由(2.82)和(2.83)定义的方向. 如果 $d\neq 0$，则 $d$ 是下降可行方向，而且 $d=0$ 的充要条件是 $x$ 为 Kuhn-Tucker 点.

计算步骤：

(1) 给定初始可行点 $x^{(1)}$，允许误差 $\varepsilon>0$，置 $k=1$.

(2) 从 $x^{(k)}$ 中选择 $m$ 个大分量，它们的下标集记作 $J_k$，$A$ 的第 $j$ 列记作 $P_j$，令 $B$ 是由 $\{P_j \mid j\in J_k\}$ 构成的 $m$ 阶矩阵，$N$ 是由 $\{P_j \mid j\notin J_k\}$ 构成的 $m\times(n-m)$ 矩阵. 由(2.81)求出 $\gamma(x_N)$，并由(2.82)和(2.83)求出 $d_N^{(k)}$ 和 $d_B^{(k)}$，从而得到搜索方向 $d^{(k)}$.

(3) 若 $\|d^{(k)}\|\leqslant\varepsilon$，则停止计算，得到点 $x^{(k)}$；否则，进行(4).

(4) 由(2.85)求 $\lambda_{\max}$，从 $x^{(k)}$ 出发，沿 $d^{(k)}$ 搜索：

$$\begin{aligned} &\min \quad f(x^{(k)} + \lambda d^{(k)}); \\ &\text{s.t.} \quad 0 \leqslant \lambda \leqslant \lambda_{\max}. \end{aligned}$$

得到最优解 $\lambda_k$.

(5) 令 $x^{(k+1)}=x^{(k)}+\lambda_k d^{(k)}$，$k:=k+1$，转(2).

现举例说明上述计算过程.

**例 2.3.9** 用 Wolfe 既约梯度法求解下列问题：

$$\begin{aligned}
\min \quad & 2x_1^2 + x_2^2; \\
\text{s.t.} \quad & x_1 - x_2 + x_3 = 2, \\
& -2x_1 + x_2 + x_4 = 1, \\
& x_j \geqslant 0, j = 1,2,3,4.
\end{aligned}$$

**解** 取初始可行点 $\boldsymbol{x}^{(1)}=(1,3,4,0)^T$.

第 1 次迭代：$J_1=\{2,3\}$，$\nabla f(\boldsymbol{x}^{(1)})=(4,6,0,0)^T$

$$\boldsymbol{x}_B=\begin{bmatrix}x_2\\x_3\end{bmatrix}=\begin{bmatrix}3\\4\end{bmatrix},\ \boldsymbol{x}_N=\begin{bmatrix}x_1\\x_4\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}.$$

把约束方程的系数矩阵 $\boldsymbol{A}$ 分解成 $\boldsymbol{B}$ 和 $\boldsymbol{N}$：

$$\boldsymbol{B}=\begin{bmatrix}-1&1\\1&0\end{bmatrix},\ \boldsymbol{N}=\begin{bmatrix}1&0\\-2&1\end{bmatrix},\ \boldsymbol{B}^{-1}=\begin{bmatrix}0&1\\1&1\end{bmatrix}.$$

根据(2.81)至(2.84)，求得既约梯度及搜索方向为

$$\gamma(\boldsymbol{x}_N^{(1)})=\begin{bmatrix}16\\-6\end{bmatrix},$$

$$\boldsymbol{d}_N^{(1)}=\begin{bmatrix}d_1^{(1)}\\d_4^{(1)}\end{bmatrix}=\begin{bmatrix}-16\\6\end{bmatrix},\quad \boldsymbol{d}_B^{(1)}=\begin{bmatrix}d_2^{(1)}\\d_3^{(1)}\end{bmatrix}=\begin{bmatrix}-38\\-22\end{bmatrix},$$

$$\boldsymbol{d}^{(1)}=(-16,\ -38,\ -22,\ 6)^T.$$

沿方向 $\boldsymbol{d}^{(1)}$ 的步长上限 $\lambda_{\max}=1/16$.

从 $\boldsymbol{x}^{(1)}$ 出发，沿 $\boldsymbol{d}^{(1)}$ 搜索，即求解下列问题：

$$\begin{aligned}
\min \quad & 2(1-16\lambda)^2+(3-38\lambda)^2; \\
\text{s.t.} \quad & 0\leqslant\lambda\leqslant\frac{1}{16}.
\end{aligned}$$

解得 $\lambda_1=1/16$.

经第 1 次迭代，得到点

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\lambda_1\boldsymbol{d}^{(1)}=\left(0,\frac{5}{8},\frac{21}{8},\frac{3}{8}\right)^T.$$

第 2 次迭代：从 $\boldsymbol{x}^{(2)}$ 出发. 这时下标集 $J_2=\{2,3\}$，即基变量为 $x_2, x_3$. 根据第 1 次迭代结果

$$x_B^{(2)} = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{5}{8} \\ \frac{21}{8} \end{bmatrix}, x_N^{(2)} = \begin{bmatrix} x_1 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{3}{8} \end{bmatrix},$$

既约梯度

$$\gamma(x_N^{(2)}) = \begin{bmatrix} \frac{5}{2} \\ -\frac{5}{4} \end{bmatrix}.$$

根据公式(2.82)至(2.84),计算得到

$$d_N^{(2)} = \begin{bmatrix} d_1^{(2)} \\ d_4^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{5}{4} \end{bmatrix}, \ d_B^{(2)} = \begin{bmatrix} d_2^{(2)} \\ d_3^{(2)} \end{bmatrix} = \begin{bmatrix} -\frac{5}{4} \\ -\frac{5}{4} \end{bmatrix},$$

$$d^{(2)} = \left(0, -\frac{5}{4}, -\frac{5}{4}, \frac{5}{4}\right)^{\mathrm{T}}.$$

从 $x^{(2)}$ 出发沿方向 $d^{(2)}$ 的步长上限 $\lambda_{\max}=1/2$. 经这次迭代得到

$$x^{(3)} = x^{(2)} + \lambda_2 d^{(2)} = (0,0,2,1)^{\mathrm{T}}.$$

在进行第 3 次迭代时,发现搜索方向 $d^{(3)}=(0,0,0,0)^{\mathrm{T}}$. 根据定理 2.3.8,现行点 $x^{(3)}=(0,0,2,1)^{\mathrm{T}}$ 是 Kuhn-Tucker 点. 由于本例是凸规划,因此 $x^{(3)}$ 是整体最优解.

### 2 广义既约梯度法

Abadie 和 Carpentier 把 Wolfe 既约梯度法推广到具有非线性约束的情形,给出**广义既约梯度法**(generalized reduced gradient method),简称为 GRG 算法. 原来的既约梯度法则简称为 RG 算法.

考虑非线性规划问题:

$$\begin{aligned} \min \quad & f(x); \\ \text{s.t.} \quad & h(x) = 0, \qquad (2.86) \\ & l \leqslant x \leqslant u. \end{aligned}$$

其中 $\boldsymbol{h}(\boldsymbol{x})=(h_1(\boldsymbol{x}),\cdots,h_m(\boldsymbol{x}))^{\mathrm{T}}$,

$$\boldsymbol{l}=(l_1,\cdots,l_n)^{\mathrm{T}},\ \boldsymbol{u}=(u_1,\cdots,u_n)^{\mathrm{T}},$$

$f,h_j(j=1,\cdots,m)$是连续可微函数,$\boldsymbol{x}\in E^n$,$m\leqslant n$.

类似于 RG 算法,把变量区分为基变量和非基变量,它们组成的向量分别用 $\boldsymbol{x}_B$ 和 $\boldsymbol{x}_N$ 表示. 相应地,把 $\boldsymbol{h}(\boldsymbol{x})$的 Jacobi 矩阵

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \frac{\partial h_1}{\partial x_1},\cdots,\frac{\partial h_1}{\partial x_n} \\ \cdots\cdots\cdots\cdots \\ \frac{\partial h_m}{\partial x_1},\cdots,\frac{\partial h_m}{\partial x_n} \end{bmatrix} \tag{2.87}$$

分解成

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}=\left(\frac{\partial h}{\partial x_B},\frac{\partial h}{\partial x_N}\right).$$

这里假设前 $m$ 个分量是基变量,并假设矩阵$\frac{\partial h}{\partial x_B}$非奇异. 这样,$\boldsymbol{x}_B$ 可以用 $\boldsymbol{x}_N$ 表示,从而把目标函数化成只是 $\boldsymbol{x}_N$ 的函数,即

$$f(\boldsymbol{x}_B(\boldsymbol{x}_N),\boldsymbol{x}_N)=F(\boldsymbol{x}_N).$$

既约梯度

$$\gamma(\boldsymbol{x}_N)\triangleq\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}_N}\triangleq\nabla_{x_N}f-\left[\left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_B}\right)^{-1}\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}_N}\right]^{\mathrm{T}}\nabla_{x_B}f, \tag{2.88}$$

其中

$$\frac{df}{d\boldsymbol{x}_N}=\left(\frac{df}{dx_{m+1}},\cdots,\frac{df}{dx_n}\right)^{\mathrm{T}}. \tag{2.89}$$

下面说明怎样利用既约梯度解决搜索方向问题. 对应非基变量 $\boldsymbol{x}_N$,定义一个向量 $\boldsymbol{d}_N^{(k)}$,使它的分量满足

$$\boldsymbol{d}_{N_j}^{(k)}=\begin{cases} 0, & \text{当 } x_{N_j}^{(k)}=l_{N_j} \text{ 且 } \gamma_j(\boldsymbol{x}_N^{(k)})>0, \\ & \text{或当 } x_{N_j}^{(k)}=u_{N_j} \text{ 且 } \gamma_j(\boldsymbol{x}_N^{(k)})<0; \\ -\gamma_j(\boldsymbol{x}_N^{(k)}), & \text{其它}. \end{cases} \tag{2.90}$$

其中 $d_{N_j}^{(k)}$是 $\boldsymbol{d}_N^{(k)}$ 的第 $j$ 个分量，$x_{N_j}^{(k)}$是 $\boldsymbol{x}_N^{(k)}$ 的第 $j$ 个分量，$l_{N_j}$和 $u_{N_j}$ 分别是非基变量 $x_{N_j}$的下界和上界.

由于 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{0}$ 是非线性方程组，不能像线性约束那样求出 $\boldsymbol{d}_B^{(k)}$ 的表达式. 为从 $\boldsymbol{x}^{(k)}$出发求出使目标函数值下降的可行点，在定义 $\boldsymbol{d}_N^{(k)}$ 以后，取适当的步长 $\lambda$，令 $\hat{\boldsymbol{x}}_N=\boldsymbol{x}_N^{(k)}+\lambda\boldsymbol{d}_N^{(k)}$，且使得

$$\boldsymbol{l}_N\leqslant\hat{\boldsymbol{x}}_N\leqslant\boldsymbol{u}_N.$$

再求解非线性方程组

$$\boldsymbol{h}(\boldsymbol{y},\hat{\boldsymbol{x}}_N)=\boldsymbol{0},\tag{2.91}$$

得到 $\hat{\boldsymbol{y}}$ . 若满足

$$f(\hat{\boldsymbol{y}},\boldsymbol{x}_N^{(k)})<f(\boldsymbol{x}_B^{(k)},\boldsymbol{x}_N^{(k)}),\tag{2.92}$$

并且

$$\boldsymbol{l}_B\leqslant\hat{\boldsymbol{y}}\leqslant\boldsymbol{u}_B,\tag{2.93}$$

则得到新的可行点$(\hat{\boldsymbol{y}},\hat{\boldsymbol{x}}_N)$；若 $\hat{\boldsymbol{y}}$ 不满足(2.92)和(2.93)，则减小步长 $\lambda$,重复以上过程.

计算步骤：

(1) 给定初始可行点 $\boldsymbol{x}^{(1)}$，允许误差 $\varepsilon_1,\varepsilon_2>0$，正整数 $J$，置 $k=1$.

(2) 将 $\boldsymbol{x}^{(k)}$分解成基变量和非基变量：$(\boldsymbol{x}_B^{(k)},\boldsymbol{x}_N^{(k)})$. 按照(2.88)计算既约梯度 $\boldsymbol{r}(\boldsymbol{x}_N)$，依据(2.90)求得方向 $\boldsymbol{d}_N^{(k)}$.

(3) 若 $\|\boldsymbol{d}_N^{(k)}\|<\varepsilon_1$，则停止计算，得到 $\boldsymbol{x}^{(k)}$；否则，进行(4).

(4) 取 $\lambda>0$，令 $\hat{\boldsymbol{x}}_N=\boldsymbol{x}_N^{(k)}+\lambda\boldsymbol{d}_N^{(k)}$. 若 $\boldsymbol{l}_N\leqslant\hat{\boldsymbol{x}}_N\leqslant\boldsymbol{u}_N$，则进行(5)；否则，以$\frac{1}{2}\lambda$代替 $\lambda$，再求 $\hat{\boldsymbol{x}}_N$，直至满足 $\boldsymbol{l}_N\leqslant\hat{\boldsymbol{x}}_N\leqslant\boldsymbol{u}_N$，进行(5).

(5) 求解非线性方程组(2.91). 采用牛顿法求解：

令 $\boldsymbol{y}^{(1)}=\boldsymbol{x}_B^{(k)}$，$j=1$，进行下列步骤：

1) 令

$$y^{(j+1)} = y^{(j)} - \left[\frac{\partial h(y^{(j)}, \hat{x}_N)}{\partial x_B}\right]^{-1} h(y^{(j)}, \hat{x}_N)$$

若 $f(y^{(j+1)}, \hat{x}_N) < f(x^{(k)})$，$l_B \leqslant y^{(j+1)} \leqslant u_B$，并且 $\| h(y^{(j+1)}, \hat{x}_N) \| < \varepsilon_2$，则转(6)；否则，进行 2)。

2) 若 $j=J$，则以 $\frac{1}{2}\lambda$ 代替 $\lambda$，令 $\hat{x}_N = x_N^{(k)} + \lambda d_N^{(k)}$，$y^{(1)} = x_B^{(k)}$，置 $j=1$，返回 1)；否则，置 $j := j+1$，返回 1)。

(6) 令 $x^{(k+1)} = (y^{(j+1)}, \hat{x}_N)$，置 $k := k+1$，返回(2)。

为了减少计算量，在解非线性方程组时，可用 $\left[\frac{\partial h(x^{(k)})}{\partial x_B}\right]^{-1}$ 近似取代 $\left[\frac{\partial h(y^{(j)}, \hat{x}_N)}{\partial x_B}\right]^{-1}$，由于前者在求既约梯度时已经计算，这样做并不增加工作量。

### 2.3.4 Frank-Wolfe 法

Frank 和 Wolfe 于 1956 年提出求解线性约束问题的一种算法。考虑非线性规划问题：

$$\left.\begin{array}{ll} \min & f(x); \\ \text{s.t.} & Ax = b, \\ & x \geqslant 0. \end{array}\right\} \tag{2.94}$$

其中 $A$ 是 $m \times n$ 矩阵，秩为 $m$。$b$ 是 $m$ 维列向量。$f(x)$ 是连续可微函数，$x \in E^n$。问题的可行域为

$$S = \{x \mid Ax = b, x \geqslant 0\}.$$

Frank-Wolfe 算法的基本思想是，在每次迭代中，将目标函数 $f(x)$ 线性化，通过解线性规划求得下降可行方向，进而沿此方向在可行域内作一维搜索。

假设 $x^{(k)}$ 是已知的可行点。在 $x^{(k)}$ 将 $f(x)$ 展开，用线性函数近似 $f(x)$，于是(2.94)转化为求解下列线性规划：

$$\left.\begin{array}{ll}\min & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{x};\\ \text{s.t.} & \boldsymbol{x}\in S.\end{array}\right\}$$

设存在最优解 $\boldsymbol{y}^{(k)}\in S$，则解线性规划的结果必为下列两种情形之一：

(1) 如果

$$\nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)})=0, \tag{2.95}$$

则 $\boldsymbol{x}^{(k)}$ 为(2.94)的 Kuhn-Tucker 点；

(2) 如果 $\nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)})\neq 0$，那么 $\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)}$ 为下降可行方向. 这时，从 $\boldsymbol{x}^{(k)}$ 出发，沿此方向作一维搜索.

计算步骤：

(1) 给定初始可行点 $\boldsymbol{x}^{(1)}$，允许误差 $\varepsilon>0$，置 $k=1$.

(2) 求解线性规划问题

$$\begin{array}{ll}\min & \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{x};\\ \text{s.t.} & \boldsymbol{x}\in S.\end{array}$$

得到线性规划的最优解 $\boldsymbol{y}^{(k)}$.

(3) 若 $|\nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)})|\leqslant\varepsilon$，则停止计算，得到点 $\boldsymbol{x}^{(k)}$；否则，进行步(4).

(4) 从 $\boldsymbol{x}^{(k)}$ 出发，沿方向 $\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)}$，在连结 $\boldsymbol{x}^{(k)}$ 和 $\boldsymbol{y}^{(k)}$ 的线段上搜索：

$$\begin{array}{ll}\min & f(\boldsymbol{x}^{(k)}+\lambda(\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)}));\\ \text{s.t.} & 0\leqslant\lambda\leqslant 1.\end{array}$$

得到 $\lambda_k$.

(5) 令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\lambda_k(\boldsymbol{y}^{(k)}-\boldsymbol{x}^{(k)})$，置 $k:=k+1$，返回步(2).

若在迭代中(2.95)总不成立，便产生序列 $\{\boldsymbol{x}^{(k)}\}$. 这个序列的每个聚点一定是 Kuhn-Tucker 点.

**定理 2.3.10** 设 $f$ 是连续可微函数，下列两个条件之一成立：

(1) $S=\{\boldsymbol{x}\mid \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geqslant\boldsymbol{0}\}$ 有界；

(2) 当 $\|x\| \to +\infty$时，$f(x) \to +\infty$.

初点 $x^{(1)} \in S$，则 Frank-Wolfe 算法收敛于(2.94)的 Kuhn-Tucker 点.

Frank-Wolfe 算法是一种可行方向法. 使用这种方法，在每次迭代中，搜索方向总是指向极点，并且当迭代点接近最优解时，搜索方向与目标函数的梯度趋于正交，这样的方向并不是最好的下降方向，因此算法收敛较慢. 但是，这种方法把求解非线性规划转化为求解一系列线性规划，在某些情形下，也能收到较好计算效果，在实际应用中仍是一种有用的算法.

### 2.3.5 近似规划方法

考虑非线性规划问题：

$$\left.\begin{aligned} \min \quad & f(x); \\ \text{s.t.} \quad & g_j(x) \geqslant 0, j = 1, \cdots, m, \\ & h_j(x) = 0, j = 1, \cdots, l. \end{aligned}\right\} \qquad (2.96)$$

其中 $x \in E^n$，$f(x)$，$g_j(x)(j=1,\cdots,m)$和 $h_j(x)(j=1,\cdots,l)$均存在一阶连续偏导数.

**近似规划法**(approximation programming method)的基本思想是，将(2.96)中的目标函数 $f(x)$和约束函数 $g_j(x)(j=1,\cdots,m)$，$h_j(x)(j=1,\cdots,l)$线性化，并对变量的取值范围加以限制，从而得到线性近似规划，再用单纯形方法求解，并把其最优解作为(2.96)的解的近似. 每得到一个近似解后，再从这解出发，重复以上步骤. 这样，通过求解一系列线性规划，产生一个由线性规划最优解组成的序列. 经验表明，这样的序列往往收敛于非线性规划问题的解.

计算步骤：

(1) 给定初始可行点 $x^{(1)}$，步长限制 $\delta_j^{(1)}$，$j=1,\cdots,n$，缩小系数 $\beta \in (0,1)$，允许误差 $\varepsilon_1$，$\varepsilon_2$，置 $k=1$.

(2) 求解线性规划问题：

$$\left.\begin{array}{ll}\min & f(\boldsymbol{x}^{(k)})+\nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^{(k)});\\ \text{s.t.} & g_j(\boldsymbol{x}^{(k)})+\nabla g_j(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^{(k)})\geqslant 0, j=1,\cdots,m,\\ & h_j(\boldsymbol{x}^{(k)})+\nabla h_j(\boldsymbol{x}^{(k)})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^{(k)})=0, j=1,\cdots,l,\\ & \quad |x_j-x_j^{(k)}|\leqslant \delta_j^{(k)}, j=1,\cdots,n.\end{array}\right\}\tag{2.97}$$

得到线性规划的最优解 $\bar{x}$.

(3) 若 $\bar{x}$ 满足可行性，则令 $x^{(k+1)}=\bar{x}$，转(4)；否则，置 $\delta_j^{(k)}:=\beta\delta_j^{(k)}, j=1,\cdots,n$，返回(2).

(4) 若 $|f(\boldsymbol{x}^{(k+1)})-f(\boldsymbol{x}^{(k)})|<\varepsilon_1$，且满足 $\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\|<\varepsilon_2$，或者

$$|\delta_j^{(k)}|<\varepsilon_2, j=1,\cdots,n.$$

则点 $\boldsymbol{x}^{(k+1)}$ 为近似最优解；否则，令 $\delta_j^{(k+1)}=\delta_j^{(k)}, j=1,\cdots,n$，置 $k:=k+1$，返回(2).

用线性近似规划方法求解非线性规划问题时，应注意以下两点：

(1) 步长限制 $\delta_j$ 的选择对算法影响很大. 如果 $\delta_j$ 取值太小，则算法收敛很慢；如果 $\delta_j$ 取值太大，则线性规划的最优解有可能不是原问题的可行解，这样不得不减小 $\delta_j$，重解当前的线性规划，增加了计算量.

(2) 关于线性规划求解方法的选择. 由于(2.97)中，变量有界，因此可用关于有界变量的单纯形方法进行求解. 否则，需引入变量的非负限制，化成标准形式，再用修正单纯形方法求解.

**例 2.3.11** 用近似规划方法解下列问题：

$$\begin{aligned}
\min\quad & f(\boldsymbol{x}) \triangleq (x_1-3)^2+(x_2-3)^2;\\
\text{s.t.}\quad & g_1(\boldsymbol{x}) = 8-x_1^2-x_2^2 \geqslant 0,\\
& g_2(\boldsymbol{x}) = x_1+x_2-1 \geqslant 0,\\
& g_3(\boldsymbol{x}) = x_1 \geqslant 0,\\
& g_4(\boldsymbol{x}) = x_2 \geqslant 0.
\end{aligned}$$

下面说明求解过程.先取初始可行点 $\boldsymbol{x}^{(1)}$和初始步长 $\boldsymbol{\delta}^{(1)}$为

$$\boldsymbol{x}^{(1)} = \begin{bmatrix}1\\1\end{bmatrix},\ \boldsymbol{\delta}^{(1)} = \begin{bmatrix}2\\2\end{bmatrix}$$

在点 $\boldsymbol{x}^{(1)}$将 $f(\boldsymbol{x})$和 $g_1(\boldsymbol{x})$线性化:

$$\begin{aligned}
f(\boldsymbol{x}) &\approx f(\boldsymbol{x}^{(1)})+\nabla f(\boldsymbol{x}^{(1)})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^{(1)})\\
&=16-4x_1-4x_2,\\
g_1(\boldsymbol{x}) &\approx g_1(\boldsymbol{x}^{(1)})+\nabla g_1(\boldsymbol{x}^{(1)})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^{(1)})\\
&=10-2x_1-2x_2.
\end{aligned}$$

规定变量取值范围:$|x_1-1|\leqslant 2$,$|x_2-1|\leqslant 2$,即要求变量满足$-1\leqslant x_1\leqslant 3$,$-1\leqslant x_2\leqslant 3$.从而得到原问题的近似线性规划:

$$\begin{aligned}
\min\quad & 16-4x_1-4x_2;\\
\text{s.t.}\quad & 10-2x_1-2x_2 \geqslant 0,\\
& x_1+x_2-1 \geqslant 0,\\
& 0 \leqslant x_1 \leqslant 3,\\
& 0 \leqslant x_2 \leqslant 3.
\end{aligned}$$

用单纯形方法求得线性规划的最优解 $\bar{\boldsymbol{x}}=\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\end{bmatrix}=\begin{bmatrix}3\\2\end{bmatrix}$.

经检验,$\bar{\boldsymbol{x}}$ 不是原来问题的可行解,因此减小步长 $\boldsymbol{\delta}^{(1)}$,取 $\beta=1/4$,令

$$\boldsymbol{\delta}^{(1)} = \beta\begin{bmatrix}2\\2\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\[4pt] \frac{1}{2}\end{bmatrix}.$$

步长改变后，仍在点 $\boldsymbol{x}^{(1)}$ 处将 $f(\boldsymbol{x})$ 和 $g_1(\boldsymbol{x})$ 线性化，得到的线性规划为

$$\begin{aligned}
\min \quad & 16 - 4x_1 - 4x_2; \\
\text{s.t.} \quad & 10 - 2x_1 - 2x_2 \geqslant 0, \\
& x_1 + x_2 - 1 \geqslant 0, \\
& \frac{1}{2} \leqslant x_1 \leqslant \frac{3}{2}, \\
& \frac{1}{2} \leqslant x_2 \leqslant \frac{3}{2}.
\end{aligned}$$

用单纯形方法求得这个线性规划的最优解

$$\bar{\boldsymbol{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}.$$

经检验，$\bar{\boldsymbol{x}}$ 是非线性规划问题的可行解，因此得到 $\boldsymbol{x}^{(1)}$ 的后继点 $\boldsymbol{x}^{(2)}$，即

$$\boldsymbol{x}^{(2)} = \bar{\boldsymbol{x}} = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}.$$

再在点 $\boldsymbol{x}^{(2)}$ 将 $f(\boldsymbol{x})$，$g_1(\boldsymbol{x})$ 线性化，取定步长 $\delta_2$，构造新的线性规划，并用单纯形方法求解. 按上述方式进行，直至得到满足精度要求的近似解.

### 2.3.6 割平面法

**割平面法**(cutting plane method)是通过解一系列线性规划来求凸规划最优解的一种方法. 问题一般可表示为

$$\left.\begin{aligned}
\min \quad & f(\boldsymbol{x}) \triangleq \boldsymbol{c}\boldsymbol{x}; \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \geqslant 0, i = 1, \cdots, m.
\end{aligned}\right\} \tag{2.98}$$

其中 $c$ 是 $n$ 维行向量，$g_i(x)(i=1,\cdots,m)$是凹函数. 问题的可行域为

$$S=\{x \mid g_i(x)\geqslant 0, i=1,\cdots,m\}.$$

一般凸规划均可写成上述形式，这是因为凸规划

$$\begin{aligned}&\min \quad f(x);\\&\text{s.t.} \quad g_i(x)\geqslant 0, i=1,\cdots,m.\end{aligned}$$

等价于凸规划

$$\begin{aligned}&\min \quad z;\\&\text{s.t.} \quad z-f(x)\geqslant 0,\\&\qquad\quad g_i(x)\geqslant 0, i=1,\cdots,m.\end{aligned}$$

下面介绍的割平面法，是由 Kelley 和 Cheney-Goldstein 于 1959 年～1960 年给出的，也称之为 KCG 法.

割平面法的基本思想是，用多面集取代可行域，并在多面集上极小化目标函数 $cx$. 运用这种方法时，首先假定可行域包含在由有限个线性不等式定义的紧集 $S_1$ 中.

解线性规划问题

$$\begin{aligned}&\min \quad cx;\\&\text{s.t.} \quad x\in S_1.\end{aligned}$$

得到此问题的最优解 $\bar{x}$ 后，在 $\bar{x}$ 将(2.98)中具有最小约束函数值的约束线性化，假设

$$g_r(\bar{x}) \triangleq \min\{g_i(\bar{x}) \mid i=1,\cdots,m\},$$

则将 $g_r(x)$线性化，并把线性化约束

$$g_r(\bar{x})+\nabla g_r(\bar{x})^{\mathrm{T}}(x-\bar{x})\geqslant 0 \tag{2.99}$$

加入到确定 $S_1$ 的线性约束集中，构成一个新的线性约束集，从而确定一个新的多面集 $S_2$，再求解线性规划

$$\left.\begin{aligned}&\min \quad cx;\\&\text{s.t.} \quad x\in S_2.\end{aligned}\right\} \tag{2.100}$$

这时必有 $S_2\subset S_1$，因为条件(2.99)的引入，使得从 $S_1$ 中割去了不

满足条件(2.99)的点.

求出(2.100)的最优解后,重复上述做法,可构造出一系列多面集

$$S_k \subset S_{k-1} \subset \cdots \subset S_2 \subset S_1,$$

它们都包含可行域$S$,而且随着$k$的增大,线性规划的可行域$S_k$越来越逼近原来问题的可行域$S$.求解一系列线性规划问题,将产生由线性规划最优解组成的一个序列$\{x^{(k)}\}$.对于凸规划,若可行域$S$有界,并且$g_i(x)$连续可微,则这个序列(至少有一个子序列)收敛于凸规划的最优解.

具体计算时,由于每次迭代与上一次相比只是增加一个新的约束条件,因此可用对偶单纯形法求解线性规划.

计算步骤:

(1) 给定初始多面集$S_1=\{x|Ax \geqslant b\}$,使得$S_1 \supset S$,$S$是问题的可行域,并且使目标函数$cx$在$S_1$上有界.给定允许误差$\varepsilon>0$,置$k=1$.

(2) 求解线性规划

$$\begin{aligned} \min \quad & cx; \\ \text{s.t.} \quad & x \in S_k. \end{aligned}$$

设线性规划最优解为$x^{(k)}$.

(3) 若$g_i(x^{(k)}) \geqslant -\varepsilon, i=1,\cdots,m$.则停止计算,得到近似解$x^{(k)}$;否则,选择下标$r$,使得

$$g_r(x^{(k)}) = \min_i \ g_i(x^{(k)}).$$

令$S_{k+1}=\{x \in S_k | g_r(x^{(k)})+\nabla g_r(x^{(k)})^{\mathrm{T}}(x-x^{(k)}) \geqslant 0\}$,置$k:=k+1$,返回(2).

**例 2.3.12** 用割平面法求解下列问题:

$$\begin{aligned}\min\quad & f(\boldsymbol{x}) \triangleq -4x_1 - x_2;\\ \text{s.t.}\quad & g_1(\boldsymbol{x}) = 8 - x_1^2 - x_2^2 \geqslant 0,\\ & g_2(\boldsymbol{x}) = 2x_2 - x_1^2 \geqslant 0,\\ & g_3(\boldsymbol{x}) = 3x_1 + x_2 - 3 \geqslant 0.\end{aligned}$$

**解** 取初始多面集

$$S_1 = \{\boldsymbol{x} \mid 0 \leqslant x_1 \leqslant 3, 0 \leqslant x_2 \leqslant 3\}.$$

容易验证 $S_1$ 包含问题的可行域.

解线性规划

$$\begin{aligned}\min\quad & -4x_1 - x_2;\\ \text{s.t.}\quad & \boldsymbol{x} \in S_1.\end{aligned}$$

得到线性规划的最优解

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

$\boldsymbol{x}^{(1)}$不满足原来的约束 $g_1(\boldsymbol{x}) \geqslant 0$ 和 $g_2(\boldsymbol{x}) \geqslant 0$，由于 $g_1(\boldsymbol{x}^{(1)}) = -10$，$g_2(\boldsymbol{x}^{(1)}) = -3$. 因此计算步骤中所用指标 $r=1$，即将 $g_1(\boldsymbol{x})$ 线性化，$g_1(\boldsymbol{x}) \approx 26 - 6x_1 - 6x_2$. 令 $S_2 = \{\boldsymbol{x} \in S_1 \mid 26 - 6x_1 - 6x_2 \geqslant 0\}$，用对偶单纯形方法解线性规划问题：

$$\begin{aligned}\min\quad & -4x_1 - x_2;\\ \text{s.t.}\quad & \boldsymbol{x} \in S_2.\end{aligned}$$

得到此线性规划问题的最优解为

$$\boldsymbol{x}^{(2)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ \dfrac{4}{3} \end{bmatrix}.$$

再判断 $\boldsymbol{x}^{(2)}$是否为原问题的最优解，由于 $g_1(\boldsymbol{x}^{(2)}) = -25/9$，$g_2(\boldsymbol{x}^{(2)}) = -19/3$，因此点 $\boldsymbol{x}^{(2)}$还不是原问题的可行解. 这时 $r=2$，按算法规定，在 $\boldsymbol{x}^{(2)}$处将 $g_2(\boldsymbol{x})$线性化

$$g_2(\boldsymbol{x}) \approx 8 - 6x_1 + 2x_2,$$

令 $S_3 = \{\boldsymbol{x} \in S_2 \mid 8 - 6x_1 + 2x_2 \geqslant 0\}$.

解线性规划：

$$\begin{aligned} \min \quad & -4x_1 - x_2; \\ \text{s.t.} \quad & \boldsymbol{x} \in S_3. \end{aligned}$$

得到线性规划的最优解

$$\boldsymbol{x}^{(3)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \dfrac{25}{12} \\ \dfrac{27}{12} \end{bmatrix}.$$

目标函数值 $f(\boldsymbol{x}^{(3)}) = -127/12$. $\boldsymbol{x}^{(3)}$已经接近原来问题的最优解 $\bar{\boldsymbol{x}} = (2,2)^{\mathrm{T}}$. 如果需要得到更精确的近似解，可继续作下去.

### 2.3.7 二次规划

**二次规划**(quadratic program)是非线性规划中一种特殊情形，它的目标函数是二次实函数，约束是线性的. 由于二次规划比较简单，便于求解，而且某些非线性规划可以转化为求解一系列二次规划问题，因此二次规划算法较早引起人们的重视，成为求解非线性规划的一个重要方法. 二次规划的算法较多，下面仅列举其中几种.

**1 Lagrange 法**

考虑二次规划问题：

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}; \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned} \tag{2.101}$$

其中 $\boldsymbol{H}$ 是 $n$ 阶对称方阵，$\boldsymbol{A}$ 是 $m \times n$ 矩阵，$\boldsymbol{A}$ 的秩为 $m$，$\boldsymbol{x} \in E^n$，$\boldsymbol{b}$ 是 $m$ 维列向量. 为介绍这种方法，首先给出下列定义.

**定义 2.3.13** Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

其中 $\boldsymbol{\lambda}$ 是 Lagrange 乘子.

设 $\bar{x}$ 是(2.101)的最优解，Lagrange 乘子为 $\bar{\lambda}$. 运用 Lagrange 方法求解(2.101)，最优解的公式是：

$$\bar{x} = -Qc + R^{\mathrm{T}}b, \tag{2.102}$$

$$\bar{\lambda} = Rc - Sb, \tag{2.103}$$

其中

$$Q = H^{-1} - H^{-1}A^{\mathrm{T}}(AH^{-1}A^{\mathrm{T}})^{-1}AH^{-1},$$

$$R = (AH^{-1}A^{\mathrm{T}})^{-1}AH^{-1},$$

$$S = -(AH^{-1}A^{\mathrm{T}})^{-1}.$$

下面给出 $\bar{x}$, $\bar{\lambda}$ 的另一种表达式.

设 $x^{(k)}$ 是任何一可行解，即 $Ax^{(k)} = b$. (2.102)和(2.103)可改写为

$$\bar{x} = x^{(k)} - Qg_k, \tag{2.104}$$

$$\bar{\lambda} = Rg_k, \tag{2.105}$$

其中 $g_k = \nabla f(x^{(k)}) = Hx^{(k)} + c$.

**2 起作用集方法**

考虑具有不等式约束的二次规划问题

$$\begin{aligned} \min \quad & f(x) \triangleq \frac{1}{2}x^{\mathrm{T}}Hx + c^{\mathrm{T}}x; \\ \text{s.t.} \quad & Ax \geqslant b. \end{aligned} \tag{2.106}$$

其中 $H$ 是 $n$ 阶对称正定矩阵，$c$ 是 $n$ 维列向量，$A$ 是 $m \times n$ 矩阵，$A$ 的秩为 $m$，$b$ 是 $m$ 维列向量，$x \in E^n$.

运用起作用集方法，在每次迭代中，以已知的可行点为起点，把在该点起作用约束作为等式约束，在此约束下极小化目标函数 $f(x)$，而其余的约束暂且不管. 求得新的比较好的可行点以后，再重复以上做法.

设在第 $k$ 次迭代中，已知可行点 $x^{(k)}$，在该点起作用约束指标

集用 $I^{(k)}$表示.这时,按算法规定,需要求解等式约束问题:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & \boldsymbol{a}^i\boldsymbol{x} = b_i, i \in I^{(k)}. \end{aligned} \tag{2.107}$$

其中 $\boldsymbol{a}^i$ 是矩阵 $\boldsymbol{A}$ 的第 $i$ 行,也是在 $\boldsymbol{x}^{(k)}$处起作用约束函数的梯度.为方便起见,现将坐标原点移至 $\boldsymbol{x}^{(k)}$,令

$$\boldsymbol{\delta} = \boldsymbol{x} - \boldsymbol{x}^{(k)},$$

(2.107)转化成求校正量 $\boldsymbol{\delta}^{(k)}$的问题:

$$\left.\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{\delta} + \nabla f(\boldsymbol{x}^{(k)})^{\mathrm{T}}\boldsymbol{\delta} \\ \text{s.t.} \quad & \boldsymbol{a}^i\boldsymbol{\delta} = 0, i \in I^{(k)} \end{aligned}\right\} \tag{2.108}$$

解二次规划(2.108),求得最优解 $\boldsymbol{\delta}^{(k)}$. 如果 $\boldsymbol{x}^{(k)}+\boldsymbol{\delta}^{(k)}$是可行点,且 $\boldsymbol{\delta}^{(k)}\neq 0$,则在第 $k+1$ 次迭代中,已知点取作 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\boldsymbol{\delta}^{(k)}$;如果 $\boldsymbol{x}^{(k)}+\boldsymbol{\delta}^{(k)}$不是可行点,则令方向 $\boldsymbol{d}^{(k)}=\boldsymbol{\delta}^{(k)}$. 沿 $\boldsymbol{d}^{(k)}$方向搜索,迭代公式如下:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}, \tag{2.109}$$

$$\alpha_k = \min\{1, \hat{\alpha}_k\}, \tag{2.110}$$

$$\hat{\alpha}_k = \min\left\{\frac{b_i - \boldsymbol{a}^i\boldsymbol{x}^{(k)}}{\boldsymbol{a}^i\boldsymbol{d}^{(k)}} \,\middle|\, i \notin I^{(k)}, \boldsymbol{a}^i\boldsymbol{d}^{(k)} < 0\right\}. \tag{2.111}$$

如果

$$\alpha_k = \frac{b_p - \boldsymbol{a}^p\boldsymbol{x}^{(k)}}{\boldsymbol{a}^p\boldsymbol{d}^{(k)}} < 1,$$

则把指标 $p$ 加入 $I^{(k)}$,得到在 $\boldsymbol{x}^{(k+1)}$处的起作用约束指标集 $I^{(k+1)}$.

如果 $\boldsymbol{\delta}^{(k)}=\boldsymbol{0}$,则 $\boldsymbol{x}^{(k)}$是(2.107)的最优解.这时应判断 $\boldsymbol{x}^{(k)}$是否为(2.106)的最优解.为此,需用公式(2.105)计算起作用约束的乘子 $\lambda_i^{(k)}, i\in I^{(k)}$. 如果这些 $\lambda_i^{(k)}\geqslant 0$,则点 $\boldsymbol{x}^{(k)}$是(2.106)的 K-T 点.由于(2.106)是凸规划,因此 $\boldsymbol{x}^{(k)}$是最优解.如果存在 $q\in I^{(k)}$,使得 $\lambda_q^{(k)}<0$,则 $\boldsymbol{x}^{(k)}$不是最优解.这时把下标 $q$ 从 $I^{(k)}$中删除.如果有几

个乘子同时为负数，令 $\lambda_q^{(k)}=\min\{\lambda_i^{(k)}|i\in I^{(k)}\}$，把对应 $\lambda_q^{(k)}$ 的约束从起作用约束集中去掉，再解(2.108).

计算步骤：

(1) 给定初始可行点 $\boldsymbol{x}^{(1)}$，相应的起作用约束指标集为 $I^{(1)}$，置 $k=1$.

(2) 求解(2.108)，设其最优解为 $\boldsymbol{\delta}^{(k)}$. 若 $\boldsymbol{\delta}^{(k)}=0$，则进行(5)；否则，进行(3).

(3) 令 $\boldsymbol{d}^{(k)}=\boldsymbol{\delta}^{(k)}$，由(2.110)求 $\alpha_k$，令 $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}$，计算 $\nabla f(\boldsymbol{x}^{(k+1)})$.

(4) 若 $\alpha_k<1$，则置 $I^{(k+1)}=I^{(k)}\cup\{p\}$，$k:=k+1$，返回(2)；若 $\alpha_k=1$，则记 $\boldsymbol{x}^{(k+1)}$ 处起作用约束指标集为 $I^{(k+1)}$，置 $k:=k+1$，进行(5).

(5) 用(2.105)计算对应起作用约束的 Lagrange 乘子 $\boldsymbol{\lambda}^{(k)}$，设

$$\lambda_q^{(k)}=\min\{\lambda_i^{(k)}|i\in I^{(k)}\}.$$

若 $\lambda_q^{(k)}\geqslant 0$，则停止计算，得最优解 $\boldsymbol{x}^{(k)}$；否则，从 $I^{(k)}$ 中删除 $q$，返回(2).

### 3 Lemke 法

Lemke 算法的基本思想是，把线性规划的单纯形方法加以适当修改，再用来求二次规划的 Kuhn-Tucker 点.

考虑二次规划问题：

$$\begin{aligned}\min\quad & f(\boldsymbol{x})\triangleq\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x};\\ \text{s.t.}\quad & \boldsymbol{A}\boldsymbol{x}\geqslant\boldsymbol{b},\\ & \boldsymbol{x}\geqslant\boldsymbol{0}.\end{aligned}\tag{2.112}$$

其中 $\boldsymbol{H}$ 是 $n$ 阶对称方阵，$\boldsymbol{c}$ 是 $n$ 维列向量，$\boldsymbol{A}$ 是 $m\times n$ 矩阵，$\boldsymbol{A}$ 的秩为 $m$，$\boldsymbol{b}$ 是 $m$ 维列向量.

引入乘子 $\boldsymbol{y}$ 和 $\boldsymbol{u}$，定义 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}. \tag{2.113}$$

再引入松弛变量 $\boldsymbol{v}\geqslant 0$,使

$$\boldsymbol{A}\boldsymbol{x} - \boldsymbol{v} = \boldsymbol{b}$$

(2.112)的 Kuhn-Tucker 条件写成:

$$\begin{cases} \boldsymbol{u} - \boldsymbol{H}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{c}, \\ \boldsymbol{v} - \boldsymbol{A}\boldsymbol{x} = -\boldsymbol{b}, \\ \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = 0, \\ \boldsymbol{v}^{\mathrm{T}}\boldsymbol{y} = 0, \\ \boldsymbol{u},\boldsymbol{v},\boldsymbol{x},\boldsymbol{y} \geqslant \boldsymbol{0}. \end{cases} \tag{2.114}$$

进而写成下列形式:

$$\begin{cases} \boldsymbol{w} - \boldsymbol{M}\boldsymbol{z} = \boldsymbol{q}, \\ \boldsymbol{w},\boldsymbol{z} \geqslant \boldsymbol{0}, \end{cases} \tag{2.115}$$

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z} = 0. \tag{2.116}$$

其中

$$\boldsymbol{w} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}, \boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}, \boldsymbol{M} = \begin{bmatrix} \boldsymbol{H} & -\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}, \boldsymbol{q} = \begin{bmatrix} \boldsymbol{c} \\ -\boldsymbol{b} \end{bmatrix},$$

$\boldsymbol{w}$, $\boldsymbol{q}$, $\boldsymbol{z}$ 均为 $m+n$ 维列向量, $\boldsymbol{M}$ 是 $m+n$ 阶矩阵. (2.115)和(2.116)一起称为**线性互补问题**(linear complementary problem).

**定义 2.3.14** 设$(\boldsymbol{w},\boldsymbol{z})$是(2.115)的一个基本可行解,且每个互补变量对$(w_i,z_i)$中有一个变量是基变量,则称$(\boldsymbol{w},\boldsymbol{z})$是**互补基本可行解**(complementary basic feasible solution).

这样,求二次规划 Kuhn-Tucker 点就转化为求互补基本可行解.

现在介绍求互补基本可行解的 Lemke 法. 分两种情形:

(1) 如果 $\boldsymbol{q}\geqslant\boldsymbol{0}$,则$(\boldsymbol{w},\boldsymbol{z})=(\boldsymbol{q},\boldsymbol{0})$就是一个互补基本可行解.

(2) 如果不满足 $\boldsymbol{q}\geqslant\boldsymbol{0}$,则引入人工变量 $z_0$,令

$$w - Mz - ez_0 = q, \tag{2.117}$$

$$w, z, z_0 \geqslant 0, \tag{2.118}$$

$$w^{\mathrm{T}} z = 0, \tag{2.119}$$

其中 $e=(1,\cdots,1)^{\mathrm{T}}$ 是分量全为 1 的 $m+n$ 维列向量.

在求解(2.117)至(2.119)之前,先引入**准互补基本可行解**(almost complementary basic feasible solution)的概念.

**定义 2.3.15** 设$(w,z,z_0)$是(2.117),(2.118)和(2.119)的一个可行解,并且满足条件:

(1) $(w,z,z_0)$是(2.117)和(2.118)的一个基本可行解.

(2) 对某个 $s\in\{1,\cdots,m+n\}$,$w_s$ 和 $z_s$ 都不是基变量.

(3) $z_0$ 是基变量,每一个互补变量对$(w_i,z_i)$$(i=1,\cdots,m+n,$ $i\neq s)$中,恰有一个变量是基变量. 则称$(w,z,z_0)$为**准互补基本可行解**.

下面说明怎样实现从一个准互补基本可行解到另一个准互补基本可行解的转换,直至求出互补基本可行解.

首先,令

$$z_0 = \max\{-q_i \mid i = 1,\cdots,m+n\} = -q_s,$$

$$z = \mathbf{0}, w = q + ez_0 = q - eq_s,$$

则 $(w,z,z_0)=(q-eq_s,\mathbf{0},-q_s)$是一个准互补基本可行解,其中 $w_i(i\neq s)$和 $z_0$ 是基变量,其余变量为非基变量. 以这个解为起始解,用主元消去法求新的准互补基本可行解,力图迫使 $z_0$ 变为非基变量. 为保持可行性,选择主元时要遵守两条规则:

(1) 若 $w_i$(或 $z_i$)离基,则 $z_i$(或 $w_i$)进基;

(2) 按照单纯形方法中的最小比值规则确定离基变量.

这样就能实现从一个准互补基本可行解到另一个准互补基本可行解的转换,直至 $z_0$ 变为非基变量,得到互补基本可行解,或者得出由式(2.117)至(2.119)所定义的可行域无界的结论.

计算步骤:

(1) 若 $\boldsymbol{q}\geqslant\boldsymbol{0}$,则 $(\boldsymbol{w},\boldsymbol{z})=(\boldsymbol{q},\boldsymbol{0})$ 是互补基本可行解,停止计算;否则,用表格形式表示方程组(2.117).设

$$-q_s=\max\{-q_i\mid i=1,\cdots,m+n\},$$

取 $s$ 行为主行,$z_0$ 对应的列为主列,进行主元消去,令 $y_s=z_s$.

(2) 设在现行表中变量 $y_s$ 下面的列为 $\boldsymbol{d}_s$.若 $\boldsymbol{d}_s\leqslant 0$,则停止计算,得到(2.117)和(2.118)的可行域的极方向;否则,按最小比值规则确定指标 $r$,使

$$\frac{\bar{q}_r}{d_{rs}}=\min\left\{\frac{\bar{q}_i}{d_{is}}\middle| d_{is}>0\right\}.$$

如果 $r$ 行的基变量是 $z_0$,则转(4);否则,进行(3).

(3) 设 $r$ 行的基变量为 $w_l$ 或 $z_l(l\neq s)$,变量 $y_s$ 进基,以 $r$ 行为主行,$y_s$ 对应的列为主列,进行主元消去.如果离基变量是 $w_l$,则令 $y_s=z_l$;如果离基变量是 $z_l$,则令 $y_s=w_l$.转(2).

(4) 变量 $y_s$ 进基,$z_0$ 离基.以 $r$ 行为主行,$y_s$ 对应的列为主列,进行主元消去.得到互补基本可行解,停止计算.

### 2.3.8 罚函数法

考虑约束问题

$$\left.\begin{array}{ll}\min & f(\boldsymbol{x});\\ \text{s.t.} & g_i(\boldsymbol{x})\geqslant 0, i=1,\cdots,m,\\ & h_j(\boldsymbol{x})=0, j=1,\cdots,l.\end{array}\right\}\tag{2.120}$$

其中 $f(\boldsymbol{x}),g_i(\boldsymbol{x})(i=1,\cdots,m),h_j(\boldsymbol{x})(j=1,\cdots,l)$ 是 $E^n$ 上的连续函数.

**罚函数法**(penalty function method)的基本思想是,利用目标函数和约束函数组成辅助函数

$$F(\boldsymbol{x},\sigma)=f(\boldsymbol{x})+\sigma P(\boldsymbol{x}).\tag{2.121}$$

$F(\boldsymbol{x},\sigma)$ 具有这样的性质:当点 $\boldsymbol{x}$ 位于可行域以外时,$F(\boldsymbol{x},\sigma)$ 取值很大,而且离可行域越远其值越大;当点在可行域内时,函数 $F(\boldsymbol{x},$

$\sigma)=f(\boldsymbol{x})$. 这样，可将原来问题转化成关于辅助函数 $F(\boldsymbol{x},\sigma)$ 的无约束极小值问题：

$$\min \quad F(\boldsymbol{x},\sigma) \triangleq f(\boldsymbol{x}) + \sigma P(\boldsymbol{x}). \tag{2.122}$$

在极小化过程中，若 $\boldsymbol{x}$ 不是可行点，则辅助函数中的第 2 项 $\sigma P(\boldsymbol{x})$ 取很大的正值，其作用迫使迭代点靠近可行域. 因此，求解问题(2.122)能够得到约束问题(2.120)的近似解，而且 $\sigma$ 越大，近似程度越好. 通常将 $\sigma P(\boldsymbol{x})$ 称为**罚项**，$\sigma$ 称为**罚因子**，$F(\boldsymbol{x},\sigma)$ 称为**罚函数**.

罚函数可有不同定义方法. $P(\boldsymbol{x})$ 的一般形式为

$$P(\boldsymbol{x}) = \sum_{i=1}^{m}\Phi(g_i(\boldsymbol{x})) + \sum_{j=1}^{l}\Psi(h_j(\boldsymbol{x})). \tag{2.123}$$

$\Phi$ 和 $\Psi$ 是满足下列条件的连续函数：

当 $y\geqslant 0$ 时，$\Phi(y)=0$；

当 $y<0$ 时，$\Phi(y)>0$.

当 $y=0$ 时，$\Psi(y)=0$；

当 $y\neq 0$ 时，$\Psi(y)>0$.

函数 $\Phi$ 和 $\Psi$ 的典型取法如下：

$$\Phi = [\max\{0, -g_i(\boldsymbol{x})\}]^{\alpha}, \Psi = |h_j(\boldsymbol{x})|^{\beta},$$

其中 $\alpha\geqslant 1, \beta\geqslant 1$，均为给定常数. 通常取作 $\alpha=\beta=2$.

**例 2.3.16** 求解下列非线性规划问题：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \triangleq (x_1-1)^2 + x_2^2; \\ \text{s.t.} \quad & g(\boldsymbol{x}) \triangleq x_2 - 1 \geqslant 0. \end{aligned}$$

**解** 定义罚函数

$$\begin{aligned} F(\boldsymbol{x},\sigma) &= (x_1-1)^2 + x_2^2 + \sigma[\max\{0, -(x_2-1)\}]^2 \\ &= \begin{cases} (x_1-1)^2 + x_2^2, & \text{当 } x_2 \geqslant 1; \\ (x_1-1)^2 + x_2^2 + \sigma(x_2-1)^2, & \text{当 } x_2 < 1. \end{cases} \end{aligned}$$

下面用解析法求解问题：

$$\min \quad F(\boldsymbol{x},\sigma).$$

根据 $F(x,\sigma)$ 的定义，有

$$\frac{\partial F}{\partial x_1} = 2(x_1 - 1),$$

$$\frac{\partial F}{\partial x_2} = \begin{cases} 2x_2, & \text{当 } x_2 \geqslant 1; \\ 2x_2 + 2\sigma(x_2 - 1), & \text{当 } x_2 < 1. \end{cases}$$

令

$$\frac{\partial F}{\partial x_1} = 0, \frac{\partial F}{\partial x_2} = 0,$$

得到罚函数的极小点（原问题的近似解）

$$\bar{x}_\sigma = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \dfrac{\sigma}{1+\sigma} \end{bmatrix}.$$

令 $\sigma \to +\infty$，则

$$\bar{x}_\sigma \to \bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$\bar{x}$ 为约束问题的最优解.

实际计算中，罚因子的选择很重要. 如果 $\sigma$ 太小，则罚函数的极小点远离约束问题的最优解；如果 $\sigma$ 太大，则给计算增加困难. 一般是取一个趋向无穷大的严格递增正数列 $\{\sigma_k\}$，从 $\sigma_1$ 开始，对每个 $k$，求解无约束问题：

$$\min \quad f(x) + \sigma_k P(x). \tag{2.124}$$

得到极小点的序列 $\{\bar{x}_{\sigma_k}\}$，在适当的条件下，这个序列收敛于约束问题的最优解. 如此通过求解一系列无约束问题来获得约束问题最优解的方法称为**序列无约束极小化方法**（sequential unconstrained minimization technique），简称为 SUMT 方法.

计算步骤：

(1) 给定初始点 $x^{(0)}$，初始罚因子 $\sigma_1$，放大系数 $c>1$，允许误差 $\epsilon>0$，置 $k=1$.

(2) $x^{(k-1)}$为初点，求解无约束问题 $\min \quad f(x)+\sigma_k P(x)$，设其极小点为 $x^{(k)}$.

(3) 若 $\sigma_k P(x^{(k)})<\epsilon$，则停止计算，得到点 $x^{(k)}$；否则，令 $\sigma_{k+1}=c\sigma_k$，并置 $k:=k+1$，返回(2).

在序列无约束极小化过程中，$F(x^{(k)},\sigma_k)$和 $f(x^{(k)})$递增，$P(x^{(k)})$递减.

**引理 2.3.17** 设 $0<\sigma_k<\sigma_{k+1}$，$x^{(k)}$和 $x^{(k+1)}$分别为取罚因子 $\sigma_k$ 及 $\sigma_{k+1}$时无约束问题的极小点，则下列各式成立：

(1) $F(x^{(k)},\sigma_k)\leqslant F(x^{(k+1)},\sigma_{k+1})$，

(2) $P(x^{(k)})\geqslant P(x^{(k+1)})$，

(3) $f(x^{(k)})\leqslant f(x^{(k+1)})$.

约束问题的最优值与 $F(x^{(k)},\sigma_k)$，$f(x^{(k)})$之间有下列关系：

**引理 2.3.18** 设 $\bar{x}$ 是问题(2.120)的最优解，且对任意的 $\sigma_k>0$，由(2.121)定义的 $F(x,\sigma_k)$存在极小点 $x^{(k)}$，则对每一个 $k$，成立

$$f(\bar{x}) \geqslant F(x^{(k)},\sigma_k) \geqslant f(x^{(k)}). \qquad (2.125)$$

罚函数法，也称外点法，所产生的序列$\{x^{(k)}\}$在适当的条件下是收敛的.

**定理 2.3.19** 设(2.120)的可行域 $S$ 非空，且存在一个数 $\epsilon>0$，使得集合

$$S_\epsilon = \{x \mid g_i(x) \geqslant -\epsilon, i = 1,\cdots,m; |h_j(x)| \leqslant \epsilon, j = 1,\cdots,l\}$$

是紧集，又设$\{\sigma_k\}$是趋向无穷大的严格递增正数列，且对每个 $k$，(2.124)存在最优解 $x^{(k)}$，则$\{x^{(k)}\}$存在一个收敛子序列$\{x^{(k_j)}\}$，并且任何这样的收敛子序列的极限都是(2.120)的最优解.

### 2.3.9 障碍函数法

这种方法又称为**内点罚函数法**，迭代中总是从内点出发，并保持在可行域内部进行搜索. 这种方法适用于下列只有不等式约束

的问题：

$$\left.\begin{array}{ll}\min & f(x);\\ \text{s.t.} & g_i(x) \geqslant 0, i = 1,\cdots,m.\end{array}\right\} \tag{2.126}$$

其中 $f(x)$，$g_i(x)(i=1,\cdots,m)$是连续函数．现将可行域记作

$$S = \{x \mid g_i(x) \geqslant 0, i = 1,\cdots,m\}.$$

为了保持迭代点含于可行域内部，我们定义**障碍函数**(barrier function)

$$F(x,\gamma) = f(x) + \gamma B(x), \tag{2.127}$$

其中 $B(x)$是连续函数，当点 $x$ 趋向可行域边界时，$B(x)\to+\infty$．两种重要的形式是：

$$B(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)}, \tag{2.128}$$

及

$$B(x) = -\sum_{i=1}^{m} \log g_i(x). \tag{2.129}$$

$\gamma$ 是很小的正数．这样，当 $x$ 趋向边界时，函数 $F(x,\gamma)\to+\infty$；否则，由于 $\gamma$ 很小，则函数 $F(x,\gamma)$的取值近似于 $f(x)$．因此可通过求解下列问题得到(2.126)的近似解．

$$\left.\begin{array}{ll}\min & F(x,\gamma);\\ \text{s.t.} & x \in \text{int}S.\end{array}\right\} \tag{2.130}$$

由于 $B(x)$的作用，在可行域边界形成“围墙”，因此(2.130)的解 $\bar{x}_\gamma$ 含于可行域内部．$B(x)$的阻挡作用是自动实现的，因此从计算的观点看，(2.130)可当作无约束问题来处理．

根据障碍函数 $F(x,\gamma)$的定义，$\gamma$ 取值越小，(2.130)的最优解越接近(2.126)的最优解；但是，$\gamma$ 太小将给(2.130)的计算带来很大困难．因此，仍采取序列无约束极小化方法，取一个严格单调减且趋于零的障碍因子数列$\{\gamma_k\}$，对每一个 $k$，从内部出发，求解问题

$$\left.\begin{array}{ll}\min & F(x,\gamma_k);\\ \text{s.t.} & x \in \text{int}S.\end{array}\right\} \tag{2.131}$$

计算步骤：

(1) 给定初始内点 $x^{(0)} \in \mathrm{int}S$，允许误差 $\varepsilon>0$，初始参数 $\gamma_1$，缩小系数 $\beta \in (0,1)$，置 $k=1$.

(2) 以 $x^{(k-1)}$ 为初点，求解下列问题：

$$\begin{aligned} \min \quad & f(x) + \gamma_k B(x); \\ \text{s.t.} \quad & x \in \mathrm{int}S. \end{aligned}$$

其中 $B(x)$ 由式(2.128)定义. 设求得的极小点为 $x^{(k)}$.

(3) 若 $\gamma_k B(x^{(k)}) < \varepsilon$，则停止计算，得到点 $x^{(k)}$；否则，令 $\gamma_{k+1} = \beta\gamma_k$，置 $k:=k+1$，返回(2).

关于内点法的收敛性有下列结论.

**定理 2.3.20** 设在(2.126)中，可行域内部非空，且存在最优解，又设对每一个 $\gamma_k$，由(2.127)定义的 $F(x,\gamma_k)$ 在 $S$ 的内部存在极小点，并且障碍函数法产生的序列 $\{x^{(k)}\}$ 存在子序列收敛于 $\bar{x}$，则 $\bar{x}$ 是问题(2.126)的最优解.

### 2.3.10 外插技术

罚函数法当罚因子 $\sigma_k$ 不断增大($\gamma_k$ 不断减小)时，Hesse 矩阵会变为病态矩阵，经验表明，这种状况将减慢算法的收敛. 为加快罚函数法的收敛性，可以采取一些措施. **外插法**(extrapolation)就是为此目的提出的. 这种方法的基本思想是，利用前面若干个障碍因子 $\gamma_k$ 及相应的解 $x^{(k)}$，来预测下一个问题的解点. 具体地讲，把罚函数 $F(x,\gamma)$(对外点法，相当于 $F(x,\sigma)=F\left(x,\dfrac{1}{\gamma}\right)$)的极小点看作罚参数 $\gamma$ 的函数，记作 $x(\gamma)$. 在某些条件下，$x(\gamma)$ 在 $\gamma=0$ 附近是 $\gamma$ 的连续函数，因此用这个函数可求得原来问题的近似解. 假设对于罚因子 $\gamma_0>\gamma_1>\cdots>\gamma_t>0$ 求得相应罚函数 $F(x,\gamma_k)$ 的极小点

$$x^{(0)}, x^{(1)}, \cdots, x^{(t)}.$$

现用这些数据将 $x(\gamma)$ 表示成 $\gamma$ 的一个 $t$ 次多项式，设

$$x(\gamma)=\sum_{j=0}^{t}P_j\gamma^j, \tag{2.132}$$

其中 $P_j$ 为 $n$ 维列向量.

将 $\gamma_k$ 及相应的解 $x^{(k)}$ 代入(2.132)，得到

$$x^{(k)}=\sum_{j=0}^{t}P_j\gamma_k^j,\ k=0,1,\cdots,t. \tag{2.133}$$

由(2.133)可确定列向量 $P_j$，$j=0,1,\cdots,t$.

若在(2.132)中，令 $\gamma=0$，则得到 $x(0)=P_0$. $x(0)$就是原问题最优解的一个更好的近似. 我们也可以用 $x(0)$作为初始点，求 $F(x,\gamma_{t+1})$的极小点，从而加速了整个计算过程.

迭代公式如下：

给定 $\gamma_0>0$，$c>1$. 令 $\gamma_i=\gamma_0/c^i$，$i=1,\cdots,t$. 假设已经求得 $F(x,\gamma_i)$的极小点 $x^{(i)}$，$i=0,1,\cdots,t$. 令

$$\begin{cases}x^{(i,0)}=x^{(i)},\ i=0,1,\cdots,t;\\ x^{(i,k)}=\dfrac{c^k x^{(i,k-1)}-x^{(i-1,k-1)}}{c^k-1},\ i=1,\cdots,t;\ k=1,\cdots,i.\end{cases} \tag{2.134}$$

则“最好”的估计 $x(0)\approx x^{(t,t)}$.

### 2.3.11 乘子法

在某些情况下罚函数的 Hesse 矩阵在迭代过程中会变成病态(ill-condition)，为了克服这个缺点 Hestenes 和 Powell 于 1968 年各自独立地提出了**乘子法**(multiplier method). 下面分两种情形给出.

#### 1 等式约束情形

考虑等式约束问题：

$$\left.\begin{aligned} \min \quad & f(\boldsymbol{x}); \\ \text{s.t.} \quad & h_j(\boldsymbol{x}) = 0, j = 1,\cdots,l. \end{aligned}\right\} \tag{2.135}$$

其中 $f$, $h_j(j=1,\cdots,l)$是二次连续可微函数,$\boldsymbol{x}\in E^n$.

为介绍乘子法,先要定义增广 Lagrange 函数(乘子罚函数):

$$\begin{aligned} \Phi(\boldsymbol{x},\boldsymbol{v},\sigma) &= f(\boldsymbol{x}) - \sum_{j=1}^{l} v_j h_j(\boldsymbol{x}) + \frac{\sigma}{2}\sum_{j=1}^{l} h_j^2(\boldsymbol{x}) \\ &= f(\boldsymbol{x}) - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}) + \frac{\sigma}{2}\boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}). \end{aligned} \tag{2.136}$$

其中

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_l \end{bmatrix}, \ \boldsymbol{h}(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_l(x) \end{bmatrix}, \quad \sigma > 0.$$

$\Phi(\boldsymbol{x},\boldsymbol{v},\sigma)$ 与 Lagrange 函数的区别在于增加了罚项 $\frac{\sigma}{2}\boldsymbol{h}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})$;而与罚函数的区别在于增加了乘子项($-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x})$).这种区别使得增广 Lagrange 函数与 Lagrange 函数及罚函数具有不同的性态.对于 $\Phi(\boldsymbol{x},\boldsymbol{v},\sigma)$,如果知道最优乘子 $\boldsymbol{v}$,只要取足够大的罚因子 $\sigma$,不必趋向无穷大,就可通过极小化 $\Phi(\boldsymbol{x},\boldsymbol{v},\sigma)$求得(2.135)的局部最优解.但是,最优乘子 $\boldsymbol{v}$ 事先未知,因此先给定充分大的 $\sigma$ 和 Lagrange 乘子的初步估计 $\boldsymbol{v}$,然后在迭代过程中修改 $\boldsymbol{v}$,力图使 $\boldsymbol{v}$ 趋向最优值.

设在第 $k$ 次迭代中,Lagrange 乘子向量的估计为 $\boldsymbol{v}^{(k)}$,罚因子 $\sigma$,$\Phi(\boldsymbol{x},\boldsymbol{v},^{(k)},\sigma)$的极小点 $\boldsymbol{x}^{(k)}$.修正乘子 $\boldsymbol{v}$ 的公式为

$$v_j^{(k+1)} = v_j^{(k)} - \sigma h_j(x^{(k)}),\ j = 1,\cdots,l. \tag{2.137}$$

然后进行第 $k+1$ 次迭代,求 $\Phi(\boldsymbol{x},\boldsymbol{v}^{(k+1)},\sigma)$的无约束极小点.继续做下去,可望乘子 $\boldsymbol{v}^{(k)}\to\bar{\boldsymbol{v}}$,从而 $\boldsymbol{x}^{(k)}\to\bar{\boldsymbol{x}}$.如果$\{\boldsymbol{v}^{(k)}\}$不收敛,或者收敛太慢,则增大参数 $\sigma$,再进行迭代.收敛快慢一般用 $\|\boldsymbol{h}(\boldsymbol{x}^{(k)})\| / \|\boldsymbol{h}(\boldsymbol{x}^{(k-1)})\|$ 来衡量.

计算步骤：

(1) 给定初始点 $x^{(0)}$，乘子向量的初始估计 $v^{(1)}$，参数 $\sigma$，允许误差 $\varepsilon>0$，常数 $\alpha>1$，$\beta\in(0,1)$，置 $k=1$.

(2) 以 $x^{(k-1)}$ 为初点，解无约束问题 $\min\ \Phi(x,v^{(k)},\sigma)$，得解 $x^{(k)}$.

(3) 若 $\|h(x^{(k)})\|<\varepsilon$，则停止计算，得到点 $x^{(k)}$；否则，进行(4).

(4) 若

$$\frac{\|h(x^{(k)})\|}{\|h(x^{(k-1)})\|}\geqslant\beta,$$

则置 $\sigma:=\alpha\sigma$，转(5)；否则，进行(5).

(5) 用公式(2.137)计算 $v_j^{(k+1)}$，$j=1,\cdots,l$，置 $k:=k+1$，转(2).

**2　一般情形**

考虑问题：

$$\begin{aligned}\min\quad & f(x);\\ \text{s.t.}\quad & g_j(x)\geqslant 0,\ j=1,\cdots,m,\\ & h_j(x)=0,\ j=1,\cdots,l.\end{aligned}\tag{2.138}$$

定义增广 Lagrange 函数

$$\Phi(x,w,v,\sigma)=f(x)+\frac{1}{2\sigma}\sum_{j=1}^{m}\{[\max(0,w_j-\sigma g_j(x))]^2-w_j^2\}$$
$$-\sum_{j=1}^{l}v_jh_j(x)+\frac{\sigma}{2}\sum_{j=1}^{l}h_j^2(x).\tag{2.139}$$

迭代中，与只有等式约束情形类似，也是取定充分大的参数 $\sigma$，以及通过修正第 $k$ 次迭代中的乘子 $w^{(k)}$ 和 $v^{(k)}$，得到第 $k+1$ 次迭代中的乘子 $w^{(k+1)}$ 和 $v^{(k+1)}$. 修正公式如下：

$$w_j^{(k+1)}=\max(0,w_j^{(k)}-\sigma g_j(x^{(k)})),j=1,\cdots,m.\tag{2.140}$$

$$v_j^{(k+1)}=v_j^{(k)}-\sigma h_j(x^{(k)}),j=1,\cdots,l\tag{2.141}$$

算法与等式约束情形相同.

# 参 考 文 献


1 Bazaraa M S and C M Shetty. Nonlinear Programming: Theory and Algorithms. Wiley, New York, 1979

2 Avriel M . Nonlinear Programming: Analysis and Methods. Prentice-Hall, Inc., 1976

3 Luenberger D G. Introduction to Linear and Nonlinear Programming. Addison-Wesley, 1973

4 McCormick G P. Nonftnear Programming: Theory, Algorithms and Application. John Wiley and Sons, New York, 1983

# 3 整数规划

## 3.1 引言

**整数规划**(integer programming)是一类要求变量取整数值的**数学规划**(mathematical programming). 若在线性规划中,要求变量取整数值时,则称为**线性整数规划**(linear integer programming),若要求变量只取 0 或 1 时,则称为 **0,1 规划**(0,1—programming). 若只要求部分变量取整数值,则称为**混合整数规划**(mixed integer programming).

1954 年,G. B. Dantzig, D. R. Fulkerson 和 S. M. Johnson 在研究**推销商问题**(traveling salesman problem)时,首先提出了破子圈方法和将问题分解成几个子问题之和的思想,这是整数规划中两大类基本方法——**割平面**(cutting plane)方法和**分支定界**(branch and bound)法的萌芽.

1958 年,R. E. Gomory 创立了一般线性整数规划的割平面算法;1960 年,A. H. Land 和 A. G. Doig 首先对推销商问题提出了一个分解算法,紧接着,E. Balas 等人将其发展成一般线性整数规划的分支定界法,从而形成了独立的整数规划分支.

从**计算复杂性**(complexity)角度看,几乎所有的整数规划问题都属于困难问题. 很少有精确的**多项式算法**(polynomial algorithm). 因此,近年来,对整数规划的研究,主要考虑各种特殊规划问题的近似算法. 如推销商问题、背包问题. 选址问题等的近似算法. 从而派生了一个新的分支——**多面体组合**(polyhedral combinatorics). 这实际上,已进入了组合最优化的研究领域. 这里,只介

绍整数规划的最基本的算法.

## 3.2 例子

**例 3.2.1 背包问题**(knapsack problem)

一个背包的容积为 $v$,现有几种物品可装,物品 $j$ 的重量为 $w_j$,体积为 $v_j(j=1,2,\cdots,n)$,问如何配装,使得既不超过背包的容积,又使装的总重量最大.

设

$$x_j = \begin{cases} 1, & \text{物品 } j \text{ 被装入背包;} \\ 0, & \text{物品 } j \text{ 不装入背包.} \end{cases}$$

则问题可写成如下的形式:

$$\begin{aligned} \max \quad & \sum_{j=1}^{n} w_j x_j; \\ \text{s.t.} \quad & \sum_{j=1}^{n} v_j x_j \leqslant v, \\ & x_j \text{ 取 } 0 \text{ 或 } 1, \quad (j = 1,\cdots,n). \end{aligned}$$

**例 3.2.2 工厂设址问题**(plant location problem)

有 $n$ 个城市$\{1,2,\cdots,n\}$,需要某种物资的数量为 $d_1,d_2,\cdots,d_n$,现计划要建造 $m$ 座工厂. 假设,在城市 $j$ 建厂,规模为 $S_j$;而投资为 $F_j$. 从城市 $i$ 到城市 $j$ 的单位运价为 $C_{ij}$,问 $m$ 个工厂应设在何处,使得既能满足需要,又使总投资最省.

设

$$y_i = \begin{cases} 1, & \text{若有一个工厂建在城市 } i; \\ 0, & \text{城市 } i \text{ 不建厂.} \end{cases}$$

设 $x_{ij}$为从城市 $i$ 运往城市 $j$ 的物资总量,则问题可写成如下的形式:

$$\min \quad \left\{ \sum_{i,j} C_{ij} x_{ij} + \sum_{i} F_i y_i \right\};$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} \leqslant S_i y_i \quad (i = 1,2,\cdots,n),$$

$$\sum_{i=1}^{n} x_{ij} \geqslant d_j \quad (j = 1,2,\cdots,n),$$

$$\sum_{i=1}^{n} y_i = m,$$

$$y_i \text{ 取 } 0 \text{ 或 } 1, x_{ij} \geqslant 0 \quad (i,j = 1,2,\cdots,n).$$

**例 3.2.3** 离散变量的数学表示

设 $x$ 只能取 $\{a_1,a_2,\cdots,a_k\}$ 中的一个数值，则可表示为如下数学形式：

$$\sum_{i=1}^{k} a_i y_i = x,$$

$$\sum_{i=1}^{k} y_i = 1; \quad y_i \text{ 取 } 0 \text{ 或 } 1.$$

**例 3.2.4** 跳跃变量的数学表示

设 $x$ 的值或者为零，或者 $L \leqslant x \leqslant U$，则可表示为如下数学形式

$$x \geqslant Ly, \quad x \leqslant Uy, \quad y \text{ 取 } 0 \text{ 或 } 1$$

**例 3.2.5 加工问题**(job scheduling problem)

有 $m$ 台同类型的机床，有 $n$ 种零件在这种机床上加工. 设加工时间分别为 $a_1,a_2,\cdots,a_n$. 问如何分配，使各种机床的总加工任务相等，或尽可能的均衡.

设

$$x_{ij} = \begin{cases} 1, & \text{若 } a_j \text{ 在机床 } i \text{ 上加工;} \\ 0, & \text{若 } a_j \text{ 不在机床 } i \text{ 上加工.} \end{cases}$$

则问题可写成如下的形式：

$$
\begin{aligned}
\min \quad & x_0; \\
\text{s.t.} \quad & \sum_{j=1}^{n} a_j x_{ij} \leqslant x_0, \quad i = 1, \cdots, m, \\
& \sum_{i=1}^{m} x_{ij} = 1, \ j = 1, \cdots, n, \\
& x_{ij} \text{ 取 } 0 \text{ 或 } 1, i = 1, \cdots, m; j = 1, \cdots, n.
\end{aligned}
$$

**例 3.2.6** 推销商问题(traveling salesman problem) ·

一个推销商从他家 $A_0$ 出发,经过预先确定的村子,$A_1, \cdots, A_n$,然后回到家.假定,村子 $A_i$ 到 $A_j$ 的距离为 $d_{ij}$,问如何选定一个行走顺序:$\{A_0, A_{i_1}, A_{i_2}, \cdots, A_{i_n}, A_0\}$,经过要去的村子,并使总的行程最短.

对任意村子 $A_i$、$A_j$,引进 0,1 变量 $x_{ij}$,若所求的行走顺序中:紧跟着村子 $A_i$ 后面的是 $A_j$,则取 $x_{ij}=1$;否则,$x_{ij}=0$.此推销商问题可写成如下数学形式:

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} \sum_{j=0}^{n} d_{ij} x_{ij}; \\
\text{s.t.} \quad & \sum_{j=0}^{n} x_{ij} = 1, \quad i = 0, 1, \cdots, n, \\
& \sum_{i=0}^{n} x_{ij} = 1, \quad j = 0, 1, \cdots, n, \\
& \sum_{i \in S} \sum_{j \notin S} x_{ij} \geqslant 1, \text{对任意的非空真子集 } S \subset \{0, 1, \cdots, n\}, \\
& \sum_{i=0}^{n} \sum_{j=0}^{n} x_{ij} = n + 1, \\
& x_{ij} \text{ 取 } 0 \text{ 或 } 1, i, j = 0, 1, \cdots, n.
\end{aligned}
$$

第一、二组条件表示在所求的行走顺序中,紧接在村子 $A_i$ 的后面和前面,恰有一个村子;第三、四组条件保证行走路线恰构成一个回路.整数规划问题的一般形式是

$$(P)\quad \min \quad x_0 = f(x_1,\cdots,x_n); \tag{3.1}$$

$$\left.\begin{aligned} \text{s.t.}\quad & \sum_{j=1}^{n} a_{ij}x_j = b_i, i = 1,\cdots,m, \\ & x_1,\cdots,x_n \geqslant 0, \\ & x_1,\cdots,x_n \text{ 取整数值.} \end{aligned}\right\} \tag{3.2}$$

为了方便,以后记 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$,若 $f(x_1,\cdots,x_n) = \sum_{i=1}^{n} c_i x_i$,则称(P)为线性整数规划问题.

## 3.3 解法概述

人们对整数规划问题,常有如下想法:(1) 因为可行方案的数目常常是有限的,因此,经过比较后,总能求得最好的方案.例如背包的装法,最多有 $2^{n-1}$种方式,推销商的行走顺序,最多有 $n!$ 种.但是,这种穷举办法,实际上是不可行的.设想计算机每秒钟能比较一百万个方式,那么,要比较完 20! 种方式,大约需要 800 年;对 $2^{60}$种方式,就要 360 多个世纪.(2) 先放弃变量的整数性要求,解一个线性规划问题,然后用"四舍五入"办法求整数解.这种办法,只有在变量的取值很大时,才有成功的可能性.而当变量的取值较小时,特别是 0—1 变量时,往往不会成功.例如考虑整数规划问题:

$$\begin{aligned} \max \quad & x_0 = 3x_1 + 13x_2; \\ \text{s.t.} \quad & 2x_1 + 9x_2 \leqslant 40, \\ & 11x_1 - 8x_2 \leqslant 82, \\ & x_1,x_2 \geqslant 0, \quad x_1,x_2 \text{ 取整数值.} \end{aligned}$$

定义域为图 3.1 中多边形 $OABD$ 内的整点,放弃整数性要求后的线性规划最优解为 $x_0=58.8, x_1=9.2, x_2=2.4$,而原整数规划最优解为 $x_0=58, x_1=2, x_2=4$.实际上,(9.2,2.4)附近的 4 个整点

(9,2),(10,2),(10,3)和(9,3)都不是整数规划的可行解.

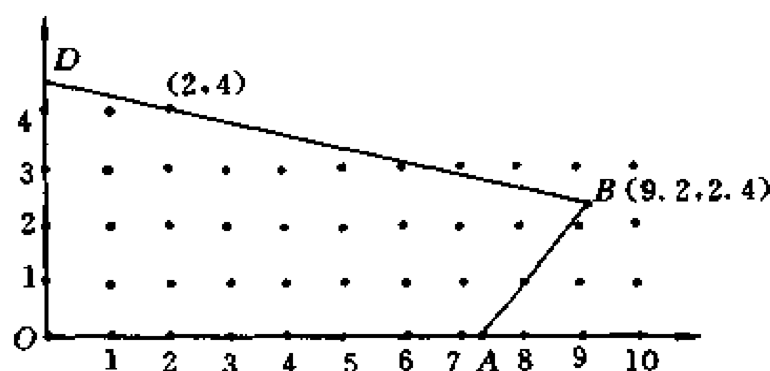



图 3.1

假如能求出多边形 $OABD$ 中的"整点凸包"(即包含 $OABD$ 中所有整点的最小多边形),如图 3.2 中的多边形 $OEFGHIJ$,则只要求此凸包上的线性规划问题的最优解,便可得到整数规划的最优解.对于一般的整数规划问题,求整点凸包是一个十分困难的问题.然而,Gomory 恰基于此想法,提出了割平面算法.

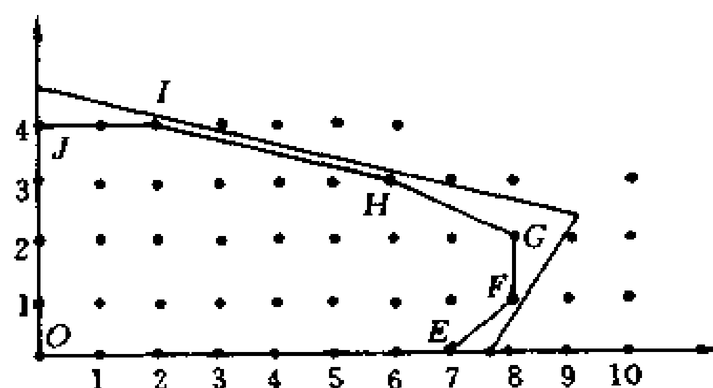



图 3.2

假如把上述整数规划问题,按照图 3.3 的树形方式,分解为 5 个定义域互不相交的整数规划子问题$\{P_1,P_2,P_3,P_4,P_5\}$之和,如图 3.4 所示,那么,$P_3$,$P_5$ 的定义域都是空集.而放弃整数性要求后,对应于 $P_1$ 的线性规划问题的最优解为(2,4),目标函数值 $x_0$

$=58$;对应于 $P_2$ 的线性规划问题的最优解为(6,3),$x_0=57$;对应于 $P_4$ 的线性规划问题的最优解为(98/11,2),$x_0=52\frac{8}{11}$. 因此,容易看出,原整数规划的最优解为(2,4). 基于这种问题分解的思想,Land, Doiog, Little, Balas 等人提出了**分支定界法**和**隐数法**(implicit enumeration method).

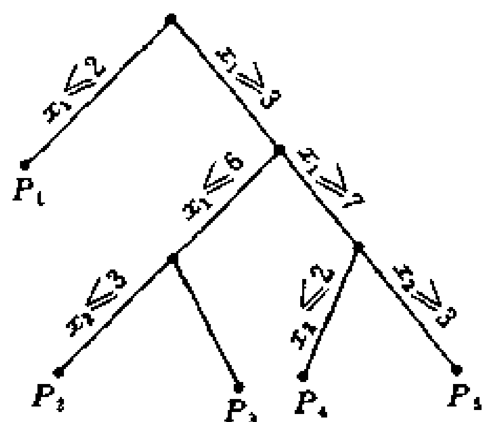



图 3.3

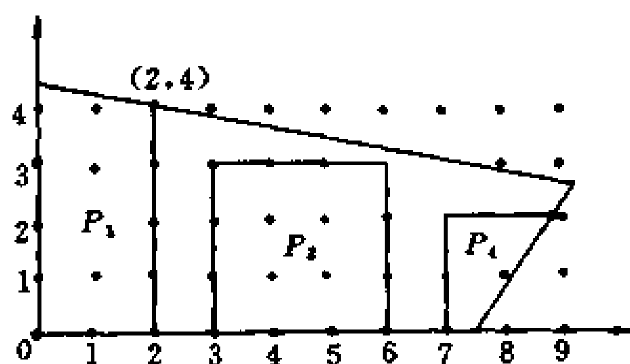



图 3.4

以上描述了目前解整数规划问题的两种基本途径. 下面叙述一般解法. 在整数规划的一般解法中,有三个基本概念.

### 3.3.1 分解(separation)

对任何整数规划问题($P$),令 $F(P)$表示($P$)的可行解集合.

若满足下述条件(3.3)和(3.4),则称问题($P$)已分解为子问题($P_1$),…,($P_q$)之和:

$$\bigcup_{i=1}^{q} F(P_i) = F(P), \tag{3.3}$$

$$F(P_i) \cap F(P_j) = \varnothing (1 \leqslant i \neq j \leqslant q). \tag{3.4}$$

常用的分解方式是"两分法".例如,若 $x_j$ 是($P$)的 0-1 变量,(即 $x_j$ 只能等于 0 或 1)则问题($P$)可以按照条件"$x_j=0$"和"$x_j=1$"分解为两个子问题之和.

### 3.3.2 松弛(relaxation)

对任何整数规划问题($P$),凡是放弃($P$)的某些约束条件,所得到的问题($\widetilde{P}$),都称为($P$)的松弛问题.对于($P$)的任何松弛问题($\widetilde{P}$),都具有如下明显的性质:

(1) 若($\widetilde{P}$)没有可行解,则($P$)也没有可行解.

(2) 对求最小值的目标函数而言,($P$)的最小值不小于($\widetilde{P}$)的最小值.

(3) 若($\widetilde{P}$)的一个最优解是($P$)的可行解,则它也是($P$)的一个最优解.

常用的松弛方式是放弃变量的整数性要求.假如($P$)是线性整数规划问题,那么松弛问题便是一个线性规划问题.

### 3.3.3 探测(fathoming)

假设按某种规则,已将问题($P$)分解为子问题($P_1$),…,($P_q$)之和,并且各($P_i$)已有对应的松弛问题($\widetilde{P}_i$).则有如下明显的性质:

(1) 若($\widetilde{P}_i$)没有可行解,则已探明了($P_i$)没有可行解.因此,可从($P$)的分解表上把它删去. (3.5)

(2) 假设当时已掌握了($P$)的一个可行解 $x^*$,它的目标函数

值为 $x_0^*$. 若松弛问题$(\tilde{P}_i)$的最小值不小于 $x_0^*$，则已探明了$(P_i)$中没有比 $\boldsymbol{x}^*$ 更好的可行解. 因此，已无需再考虑$(P_i)$，可从分解表上把它删去. (3.6)

(3) 若$(\tilde{P}_i)$的最优解是$(P_i)$的可行解，则已求得了$(P_i)$的一个最优解，因此，无需再考虑$(P_i)$了，可从表上删去. 这时，若$(P_i)$的最优解比 $\boldsymbol{x}^*$ 好，那么，替换 $\boldsymbol{x}^*$，同时也刷新 $x_0^*$ 的记录. (3.7)

(4) 假如表上各个$(\tilde{P}_i)$的目标函数最小值都不比 $x_0^*$ 小，那么，当时的记录解 $\boldsymbol{x}^*$，便是原问题$(P)$的最优解.

通常称情形(1)为**可行性探测**，(2)—(4)为**最优性探测**.

总之，求解整数规划问题$(P)$，有以下步骤. 首先，选定一种松弛方式，将$(P)$松弛成问题$(\tilde{P})$，使得较易求解. 若$(\tilde{P})$没有可行解，则$(P)$也没有可行解. 若$(\tilde{P})$的最优解是$(P)$的可行解，则已求得$(P)$的最优解. 若$(\tilde{P})$的最优解，不是$(P)$的可行解，则有两条不同的途径可走. 一是设法改进松弛问题$(\tilde{P})$，坚持继续探测$(P)$. 二是选定一种分解方式，把$(P)$分解成几个子问题之和，列表记录下来. 并且，赋予各子问题一个尽可能好的目标函数值的下界. 然后，按一定的次序，逐个进行探测. 当某个子问题已经被探明时，就从表中删去. 否则，继续对子问题进行分解.

对线性整数规划问题，不用分解技术的算法是割平面方法. 它用线性规划问题作为松弛问题. 通过逐次生成割平面条件来不断地改进松弛问题，使最后求得的松弛问题的最优解也是整数解. 利用分解技术的算法有两类：**隐数法**(implicit enumeration method)和**分支定界法**(branch and bound method). 它们通常都是用线性规划问题作为松弛问题. 不同之处，仅在于探测子问题的先后次序. 隐数法是按照子问题表中“先入后出”的原则来确定探测的先后顺序. 分支定界法是按照所赋予的下界的大小来确定探测的先后顺序. 下界小的子问题优先探测. 前者，计算程序比较简单，在计算过程中，需要保存的中间信息较少. 而后者，选取子问题时有灵

活性，因为人们可以期望在下界小的子问题中，存在（整数）最优解.

## 3.4 整数规划的一般解法

利用分解、松弛、探测技术，求解整数规划的一般过程可描述如下：

设 $\Pi$ 表示子问题表，$\boldsymbol{x}^*$ 表示$(P)$的已知的可行解，$x_0^*$ 表示 $\boldsymbol{x}^*$ 的目标函数值. 假设问题$(P)$的目标函数是求最小值.

(1) 置 $\Pi=\{(P)\}$，$\boldsymbol{x}^*=\varnothing$，$x_0^*=+\infty$.

(2) 若 $\Pi=\varnothing$，则步骤终止. $\boldsymbol{x}^*$ 便是$(P)$的最优解（若 $\boldsymbol{x}^*=\varnothing$，则$(P)$无可行解），否则，执行(3).

(3) 按"先入后出"或"下界次序"，从 $\Pi$ 中取出一个子问题，记为$(CP)$（即 $\Pi\backslash(CP)\to\Pi$).

(4) 选取$(CP)$的一个松弛问题$(\widetilde{CP})$.

(5) 选取一个"适当的"算法，求解$(\widetilde{CP})$.

(6) (3.5)探测：若通过求解$(\widetilde{CP})$，知道了$(CP)$无可行解，则转(2). 否则，执行(7).

(7) (3.6)探测：若通过求解$(\widetilde{CP})$，知道了$(CP)$中没有比 $\boldsymbol{x}^*$ 更好的解（例如$(\widetilde{CP})$的最小值$\geqslant x_0^*$），则转到(2). 否则，执行(8).

(8) (3.7)探测：若求得$(\widetilde{CP})$的最优解是$(CP)$的可行解，则转(12). 否则，执行(9).

(9) 决定是否要分解$(CP)$. 若是，则转到(11). 否则转(10).

(10) 改进$(CP)$的松弛问题$(\widetilde{CP})$（例如在原松弛问题中增加一个约束条件）. 然后，转(5).

(11) 选取一种适当的方式,分解$(CP)$. 赋予$(CP)$的各子问题尽可能好的目标函数值的下界,并将它的各子问题写入 $\Pi$ 中. 然后,转(3).

(12) 设求得$(CP)$的最优解为 $\bar{x}$,最小值为 $v(\bar{x})$. 若 $v(\bar{x})\geqslant x_0^*$,则转到(2). 若 $v(\bar{x})<x_0^*$,则 $\bar{x}\to x^*$,$v(\bar{x})\to x_0^*$,然后转(2).

### 3.4.1 分支定界法

对线性混合整数规划问题$(P)$:

$$\min \quad x_0=\sum_{j=1}^{n}c_jx_j+\sum_{k=1}^{q}d_ky_k; \tag{3.8}$$

$$\text{s.t.}\quad \left.\begin{array}{l}\sum_{j=1}^{n}a_{ij}x_j+\sum_{k=1}^{q}b_{ik}y_k=b_i,\quad i=1,\cdots,m,\\ x_1,\cdots,x_n,y_1,\cdots,y_k\geqslant 0,\\ x_0,x_1,\cdots,x_n\text{ 取整数值.}\end{array}\right\} \tag{3.9}$$

可用分支定界法求解.

计算步骤:

首先作如下的处理:取$(\tilde{P})$为放弃 $x_j$ 的整数性要求后的线性规划问题. 用单纯形方法求解$(\tilde{P})$. 若$(\tilde{P})$无可行解,则终止,这时$(P)$无可行解. 若$(\tilde{P})$无最小值,则终止,这时$(P)$无最小值. 若$(\tilde{P})$有最优解 $\tilde{x}$,最小值为 $v(\tilde{x})$. 这时,若 $\tilde{x}$ 是$(P)$的可行解,则终止,这时 $\tilde{x}$ 也是$(P)$的最优解. 若 $\tilde{x}$ 不是$(P)$的可行解,则赋予$(P)$的目标函数值下界为 $v(\tilde{x})$.

(1) 置 $\Pi=\{(P)\}$,$x^*=\varnothing$,$x_0^*=+\infty$.

(2) 若 $\Pi=\varnothing$,则终止,$x^*$ 便是$(P)$的最优解(若 $x^*=\varnothing$,则$(P)$无可行解).

(3) 从 $\Pi$ 中,取出一个使下界值最小的子问题(或按"先入后出"原则选出一个子问题),记作$(CP)$. $\Pi\backslash(CP)\to\Pi$.

(4) 对$(CP)$放弃 $x_j$ 的整数性要求，得$(\widetilde{CP})$.

(5) 用单纯形法求解$(\widetilde{CP})$，得最优解及最优目标函数值 $v(\widetilde{CP})$.

(6) (3.5)探测：若$(\widetilde{CP})$无可行解，则转(2).

(7) (3.6)探测：若$(\widetilde{CP})$的最小值 $v(\widetilde{CP})\geqslant x_0^*$，则转(2).

(8) (3.7)探测：若求得$(\widetilde{CP})$的最优解 $\tilde{x}$ 是$(CP)$的可行解，则转(12)，否则执行(9).

(9) 决定是否要分解$(CP)$，若是，则转到(11). 否则执行(10).

(10) 用"分数割平面"方法(见 3.5.4)求解$(CP)$. 并规定这一步被允许的最长计算时间(或加割平面的最多次数). 假设计算结束时，所得$(CP)$的松弛问题为$(\widetilde{CP})$(可能题没有解完，但规定的时间已到)，它的最优解为 $\tilde{x}$，最小值为 $v(\tilde{x})$(假如$(\widetilde{CP})$没有可行解，则记 $\tilde{x}=\varnothing$)，然后转(6).

(11) 按某种"启发性"的原则，选取一个整数变量 $x_j$，使它在 $\tilde{x}$ 中取非整数 $\tilde{x}_j$. 记$[\tilde{x}_j]$为不超过 $\tilde{x}_j$ 的最大整数(例如在 $\tilde{x}_1,\cdots,\tilde{x}_n$ 中，选取一个最靠近整数的分数，作为 $\tilde{x}_j$). 按条件"$x_j\leqslant[\tilde{x}_j]$"和"$x_j\geqslant[\tilde{x}_j]+1$"，将$(CP)$分解为两个子问题$(CP_1)$和$(CP_2)$，并赋予它们下界 $v(\tilde{x})$(只要再稍作努力，就可分别给出$(CP_1)$和$(CP_2)$的更好的下界). 将$(CP_1)$和$(CP_2)$，以及它们的下界记入 $\Pi$ 中，然后转(3).

(12) 若 $v(\tilde{x})<x_0^*$，则 $\tilde{x}\rightarrow x^*$，$v(\tilde{x})\rightarrow x_0^*$，然后转(2). 若 $v(\tilde{x})\geqslant x_0^*$，则转(2).

**注**：在(11)中，有一种选取分解变量 $x_j$ 的方法，称做"罚款"方法. 设用单纯形法，求得$(\widetilde{CP})$的基本最优解为

$$x_i = a_{i0} + \sum_{j\in R} a_{ij}t_j, \qquad i = 0,1,\cdots,n.$$

其中，$t_j$ 表示非基变量，$a_{i0}\geqslant 0(i=1,\cdots,n)$，$a_{0j}\geqslant 0(j\in R)$. 设 $a_{i0}$不全是整数，这时，$(CP)$的下界为 $a_{00}$.

现在，设 $a_{i0}$不是整数，按条件 $x_i\leqslant[a_{i0}]$和 $x_i\geqslant[a_{i0}]+1$，将$(CP)$分解成$(CP_1)$和$(CP_2)$. 考虑它们下界的变化情况.

(1) 若在$(\widetilde{CP})$中，增加条件 $x_i\leqslant[a_{i0}]$，则相当于在单纯形表中增加一行

$$S_i = -r_{i0} - \sum_{j\in R} a_{ij}t_j \qquad (\text{其中 } r_{i0} = a_{i0} - \lfloor a_{i0}\rfloor).$$

应用对偶单纯形规则，变换一步后，目标函数值将增加

$$D_i = \frac{a_{0s}}{a_{is}}r_{i0}, \text{其中} \frac{a_{0s}}{a_{is}} = \min_{j\in R}\left\{\frac{a_{0j}}{a_{ij}} \mid a_{ij} > 0\right\}.$$

(2) 若在$(\widetilde{CP})$中，增加条件 $x_i\geqslant\lfloor a_{i0}\rfloor+1$，则相当于在单纯形表中增加一行

$$S_i = -1 + r_{i0} + \sum_{j\in R} a_{ij}t_j.$$

应用对偶单纯形规则，变换一步后，目标函数值将增加

$$U_i = (-1 + r_{i0})\frac{a_{0s}}{a_{is}},$$

其中，$\frac{a_{0s}}{-a_{is}} = \min_{j\in R}\left\{\frac{a_{0j}}{-a_{ij}} \mid a_{ij} < 0\right\}$.

记 $P_i=\min\{D_i,U_i\}$，称为变量 $x_i$ 的"罚款". 当选取 $x_i$ 为分解变量时，$(CP)$各子问题的下界至少可增加 $P_i$. "罚款"方法就是选取使 $P_i$ 最大的变量 $x_i$ 作为分解变量. 即选分解变量 $x_j$，使满足 $P_j=\max\{P_i \mid a_{i0}\text{不是整数}\}$.

**例 3.4.1**

$$\begin{aligned}(P)\quad \min\quad & x_0 = 7x_1 + 3x_2 + 4x_3;\\ \text{s.t.}\quad & x_1 + 2x_2 + 3x_3 \geqslant 8,\end{aligned}$$

$$3x_1 + x_2 + x_3 \geqslant 5,$$
$$x_1, x_2, x_3 \geqslant 0.$$
$$x_0, x_1, x_2, x_3 \text{ 取整数}.$$

**解** 用分支定界法计算：

放弃变量整数性要求的问题记为$(\tilde{P})$，用单纯形法求解$(\tilde{P})$，得$\tilde{x}=(0.4, 3.8, 0)$，$v(\tilde{x})=14.2$，赋与$(P)$的下界为 15.

(1) 置$\Pi=\{(P)\}$，$x^*=\varnothing$，$x_0^*=+\infty$.

(2) 因$\Pi\neq\varnothing$，转(3).

(3) 取$(P)=(CP)$，$\Pi\setminus(CP)\to\Pi$. 执行(4)～(9)后，转(11).

(11) 按条件"$x_2\leqslant 3$"，和"$x_2\geqslant 4$"，将$(P)$分解为子问题$(P_1)$，$(P_2)$之和. 赋予它们的下界为 15. 置$\Pi\cup\{(P_1),(P_2)\}\to\Pi=\{(P_1),(P_2)\}$. 然后转(3).

(3) 取$(CP)=(P_1)$，置$\Pi=\{(P_2)\}$.

(4) 置$(\tilde{P}_1)$为$(\tilde{P})$加条件 $x_2\leqslant 3$.

(5) 解$(\tilde{P}_1)$，得$x^1=(0.5, 3, 0.5)$，$v(\tilde{P}_1)=14.5$. 执行(6)～(9)后，转(11).

(11) 按条件"$x_1=0$"，"$x_1\geqslant 1$"将$(P_1)$分解为$(P_3)$，$(P_4)$之和，赋予它们的下界为 15. 置$\Pi=\{(P_3),(P_4),(P_2)\}$. 然后转(3).

(3) 取$(CP)=(P_3)$，置$\Pi=\{(P_4),(P_2)\}$.

(4) 取$(\tilde{P}_3)$为$(\tilde{P})$加条件 $x_2\leqslant 3, x_1=0$.

(5) 解$(\tilde{P}_3)$，得$x^3=(0, 3, 2)$，$v(\tilde{P}_3)=17$. 执行(6)～(8)后，转(12).

(12) 置$x^*=x^3$，$x_0^*=17$，然后转(3).

(3) 取$(CP)=(P_4)$，置$\Pi=\{(P_2)\}$.

(4) 取$(\tilde{P}_4)$为$(\tilde{P})$加条件 $x_2\leqslant 3, x_1\geqslant 1$.

(5) 解$(\tilde{P}_4)$，得$x^4=\left(1, 0, 2\frac{1}{3}\right)$，$v(\tilde{P}_4)=16\frac{1}{3}$，$v(P_4)\geqslant 17$.

(6) $(\tilde{P}_4)$有可行解，执行(7).

(7) $v(P_4)\geqslant x_0^*$，$(P_4)$探明. 转(3).

(3) 取$(CP)=(P_2)$，置 $\Pi=\varnothing$.

(4) 取$(\widetilde{P}_2)$为$(\widetilde{P})$加条件 $x_2\geqslant 4$.

(5) 解$(\widetilde{P}_2)$，得 $x^2=\left(\frac{1}{3},4,0\right)$，$v(\widetilde{P}_2)=14\frac{1}{3}$.

执行(6)～(9)后，转(11).

(11) 按条件“$x_1=0$”，“$x_1\geqslant 1$”将$(P_2)$分解为$(P_5)$与$(P_6)$之和. 赋予它们的下界为 15. 置 $\Pi=\{(P_5),(P_6)\}$. 然后转(3).

(3) 取$(CP)=(P_5)$，置 $\Pi=\{(P_6)\}$.

(4) 取$(\widetilde{P}_5)$为$(\widetilde{P})$加条件 $x_2\geqslant 4$ 和 $x_1=0$.

(5) 解$(\widetilde{P}_5)$，得 $x^5=(0,5,0)$，$v(\widetilde{P}_5)=15$.

执行(6)～(8)后，转(12).

(12) 因 $v(P_5)<x_0^*$. 故置 $x^*=x^5$，$x_0^*=15$. 然后，转(3).

(3) 取$(CP)=(P_6)$，置 $\Pi=\varnothing$. 执行(4)～(6)后，转(7).

(7) 因 $v(P_6)\geqslant x_0^*$. $(P_6)$探明. 转(2).

(2) 因 $\Pi=\varnothing$，步骤终止，$(P)$的最优解为 $x^*=x^5=(0,5,0)$，$x_0^*=15$.

算法过程如图 3.5 所示.

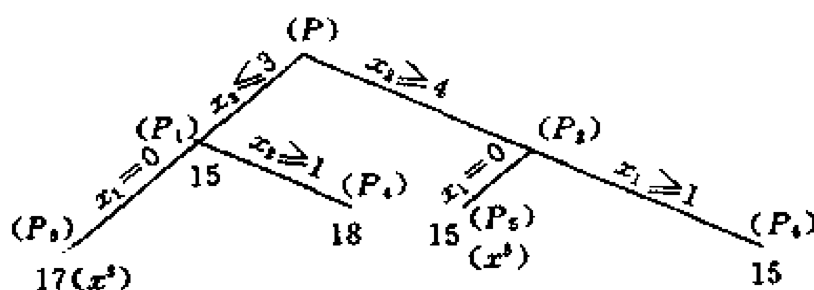



图 3.5

### 3.4.2 0-1规划的"隐数法"

考虑一类特殊的，但应用广泛的整数规划问题——线性 0-1 规划问题($P$)：

$$\min \quad x_0 = cx = \sum_{j=1}^{n} c_j x_j; \tag{3.10}$$

$$\left.\begin{aligned} \text{s.t.} \quad a_i x = \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i \quad & i = 1, \cdots, m, \\ x_j \text{ 取 } 0 \text{ 或 } 1, \quad & j = 1, \cdots, n, \end{aligned}\right\} \tag{3.11}$$

其中 $c_j, a_{ij}, b_i$ 都是整数；$a_i = (a_{i1}, a_{i2}, \cdots, a_{in})$.

不妨设所有的 $c_j \geqslant 0$. 因为若有某 $c_j < 0$，则作变换 $x_j' = 1 - x_j$ 后，目标函数中 $x_j'$ 的系数就变为 $-c_j > 0$. 也不妨设 $0 \leqslant c_1 \leqslant c_2 \leqslant \cdots \leqslant c_n$.

对任意的变量 $x_j$，若($P$)按条件 $x_j = 0$ 和 $x_j = 1$ 分为两个子问题($P_1$)，($P_2$)之和，则记($P_1$)为$\{-j\}$，记($P_2$)为$\{+j\}$. 在此子问题中，$x_j$ 称为固定变量，其余变量称为自由变量. 一般地，记 $x_i, x_j, x_k, \cdots$ 等依次已固定为 1,0,1,…的子问题为$\{+i, -j, +k, \cdots\}$. 所有已取固定值的变量称为固定变量，未取定值的变量称为自由变量. 通常用 $\sigma$ 表示固定变量指标(指带+，－号)的某一个排列. 用$\{\sigma\}$表示对应的子问题. ($P$)也记为$\{\varnothing\}$. 对任何子问题$\{\sigma\}$，定义它的松弛问题$\{\tilde{\sigma}\}$为放弃所有不等式条件 $a_i x \geqslant b_i$ 后的问题. 即为 $\min\left\{ \sum_{j \notin \sigma} c_j x_j \mid x_j \text{ 取 } 0 \text{ 或 } 1, x_j \text{ 为自由变量} \right\}$. 在松弛问题$\{\tilde{\sigma}\}$中，所有的自由变量都取 0 的解，记为 $\sigma_0$. 由于 $c_j \geqslant 0 (j = 1, \cdots, n)$，故$\{\tilde{\sigma}\}$的最优解显然是 $\sigma_0$. 如上所述，可将($P$)分解成如图 3.6 的树枝形式.

"隐数法"是沿着分解树的各个树枝，从左到右，依次探测各子问题. 下面介绍一些探测规则.

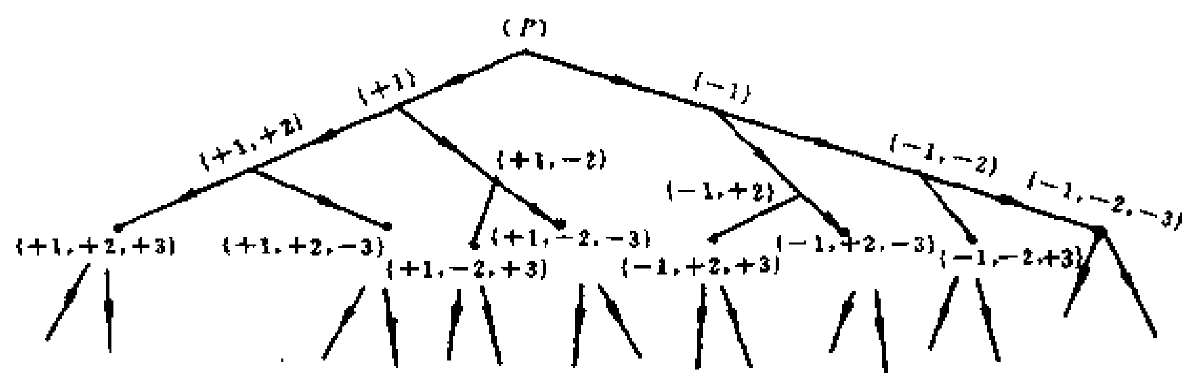



图 3.6

假设已掌握$(P)$的一个可行解 $\boldsymbol{x}^*$，它的目标函数值为 $x_0^*=\boldsymbol{c}\boldsymbol{x}^*$. 现在考虑(P)的任一子问题$\{\sigma\}$. 设 $x_j$ 是$\{\sigma\}$中下标最小的自由变量.

(1) 若 $\boldsymbol{c}\boldsymbol{\sigma}_0\geqslant x_0^*$，则$\{\sigma\}$中没有比 $\boldsymbol{x}^*$ 更好的可行解. 因此$\{\sigma\}$已探明.

(2) 若 $\boldsymbol{c}\boldsymbol{\sigma}_0<x_0^*$，且 $\sigma_0$ 是$(P)$的可行解，则置 $\boldsymbol{x}^*=\boldsymbol{\sigma}_0, x_0^*=\boldsymbol{c}\boldsymbol{\sigma}$. $\{\sigma\}$已探明.

(3) 若 $\boldsymbol{c}\boldsymbol{\sigma}_0<x_0^*$，$\sigma_0$ 不是$\{\sigma\}$的可行解，但是 $\boldsymbol{c}\boldsymbol{\sigma}_0+c_j\geqslant x_0^*$，则显然$\{\sigma\}$中也没有比 $\boldsymbol{x}^*$ 更好的可行解. $\{\sigma\}$已探明.

(4) 设自由变量为 $x_{j1}, x_{j2}, \cdots, x_{jk}$，满足 $\boldsymbol{c}\boldsymbol{\sigma}_0+c_{j1}\leqslant \boldsymbol{c}\boldsymbol{\sigma}_0+c_{j2}\leqslant\cdots\leqslant \boldsymbol{c}\boldsymbol{\sigma}_0+c_{jr}<x_0^*\leqslant \boldsymbol{c}\boldsymbol{\sigma}_0+c_{jr+1}\leqslant\cdots\leqslant \boldsymbol{c}\boldsymbol{\sigma}_0+c_{jk}$. 记 $J=\{j_1, j_2, \cdots, j_r\}$. 称 $J$ 为可选集.

令 $s_i=\boldsymbol{a}_i\boldsymbol{\sigma}_0-b_i(i=1,2,\cdots,m)$. 若所有的 $s_i\geqslant 0$，则 $\boldsymbol{\sigma}_0$ 为$(P)$的可行解. 相反

令 $I=\{i\,|\,s_i<0\}$，

$$J_i=\{j\,|\,j\in J, a_{ij}>0\},\quad i\in I,$$

$$t_i=\sum_{j\in J_i}a_{ij}\qquad i\in I,$$

容易验证

(1) 若有某 $i\in I$，使得 $s_i+t_i<0$，则 $\{\sigma\}$ 中无可行解.

(2) 若有某 $i\in I, j\in J\backslash J_i$，使得 $s_i+t_i+a_{ij}<0$，则 $\{\sigma\}$ 中没有 $x_j=1$ 的可行解. 因此可从 $J$ 中去掉此指标 $j$，重新定义各 $J_i$ 和 $t_i$.

利用上述简单的探测方式，下面介绍一个初等方法. 它在计算过程中只需用加减法.

计算步骤：

任选 $(P)$ 的一可行解作为初始的记录解 $\boldsymbol{x}^*$，置 $x_0^*=\boldsymbol{cx}^*$（假如找不到可行解，则置 $\boldsymbol{x}^*=\varnothing, x_0^*=+\infty$），置 $\{\sigma\}=\{\varnothing\}$，$\boldsymbol{\sigma}_0=(0,\cdots,0)$.

(1) 若 $\boldsymbol{c\sigma}_0\geqslant x_0^*$，则转(8). 否则计算 $s_i=\boldsymbol{a}_i\boldsymbol{\sigma}_0-b_i(i=1,\cdots,m)$，然后执行(2).

(2) 若 $s_i\geqslant 0(i=1,\cdots,m)$，即 $\boldsymbol{\sigma}_0$ 是 $(P)$ 的可行解，则置 $\boldsymbol{x}^*=\boldsymbol{\sigma}_0, x_0^*=\boldsymbol{c\sigma}_0$. 然后转(8). 否则，置 $I=\{i\mid s_i<0\}$. 设自由变量为 $x_{j_1}, x_{j_2},\cdots,x_{j_k}$ $(j_1<j_2<\cdots<j_k)$. 此时，若 $\boldsymbol{c\sigma}_0+c_{j_1}\geqslant x_0^*$，则转(8)；否则执行(3).

(3) 计算 $\boldsymbol{c\sigma}_0+c_{j_t}, t=1,2,\cdots,k$. 置 $J=\{j_t\mid \boldsymbol{c\sigma}_0+c_{j_t}<x_0^*, 1\leqslant t\leqslant k\}$.

(4) 置 $J_i=\{j\mid j\in J, a_{ij}>0\}, i\in I$,

$$t_i=\sum_{j\in J_i}a_{ij},\qquad\qquad i\in I.$$

(5) 若有某个 $i\in I$，使得 $s_i+t_i<0$，则转(8)；否则执行(6).

(6) 考察每一个指标 $j\in J$，若存在某个 $i\in I$，使得 $j\notin J_i$，而 $s_i+t_i+a_{ij}<0$，则 $J\backslash\{j\}\to J$. 此时，若 $J=\varnothing$，则转(8). 若 $J\neq\varnothing$，则转(4).

(7) 选取 $q=\min\{j\mid j\in J\}$. $\{\sigma,+q\}\to\{\sigma\}$，然后转(1).

(8) 若 $\{\sigma\}$ 中的固定变量都是固定取 0 的变量，则步骤终止，已探测完所有的子问题. 此时的 $\boldsymbol{x}^*$，若不是 $\varnothing$，便是 $(P)$ 的最优解；若是 $\varnothing$，则 $(P)$ 无可行解. 若 $\{\sigma\}$ 中的固定变量，不全是取 0 的

变量，设$\{\sigma\}=\{\cdots,+r,-s,\cdots,-t\}$，即设$x_r$是$\{\sigma\}$中最后被固定为1的变量，则$\{\cdots,-r\}\to\{\sigma\}$. 然后转到步骤(1).

**注**：在(7)中，$q$的取法是从"最优性"角度来选择. 也可从"可行性"角度来选取. 例如，对每一指标$j\notin\sigma$，令

$$D_{ij}=\min(s_i+a_{ij},0),$$

$$D_j=\sum_{i=1}^{m}D_{ij}.$$

设
$$D_f=\max_{j\notin\sigma}D_j,$$

则可取$q=f$.

**例 3.4.2**

$$\begin{aligned}\min\quad & x_0=x_1+x_2+4x_3+6x_4+7x_5;\\ \text{s.t.}\quad & 3x_1-2x_2+6x_3+2x_4+4x_5\geqslant 4,\\ & -5x_1+x_2+x_3+x_4+6x_5\geqslant 2,\\ & x_1,\cdots,x_5\text{ 取 0 或 1}.\end{aligned}$$

**解** 用隐数法计算步骤计算：

因$\boldsymbol{x}=(0,0,0,0,1)$是一可行解，取$x_0^*=7$. 置$\{\sigma\}=\{\varnothing\}$，$\boldsymbol{\sigma}_0=(0\ 0\ 0\ 0\ 0)$.

(1) 计算$(s_1,s_2)=(-4,-2)$.

(2) $I=\{1,2\}$.

(3) 根据条件$c\sigma_0+c_j<x_0^*$，得$J=\{1,2,3,4\}$.

(4) $J_1=\{1,3,4\}$，$J_2=\{2,3,4\}$，$t_1=3+6+2=11$，
$t_2=1+1+1=3$.

(5) $s_1+t_1=7\qquad s_2+t_2=1$.

(6) $s_2+t_2+a_{21}=1-5=-4\Rightarrow J=\{2,3,4\}$.

(7) $q=2$，$\{\sigma\}=\{+2\}$.

(1) $(s_1,s_2)=(-6,-1)$；

(2) $I=\{1,2\}$.

(3) $J=\{1,3\}$.

(4) $J_1=\{1,3\}$，$J_2=\{3\}$，$t_1=3+6=9$，$t_2=1$.

(5) $s_1+t_1=3$，$s_2+t_2=0$.

(6) $s_2+t_2+a_{21}=-5 \Rightarrow J=\{3\}$.
(7) $q=3$, $\{\sigma\}=\{+2,+3\}$.
(1) $(s_1,s_2)=(0,0)$.
(2) 置 $x^*=(0,1,1,0,0)$, $x_0^*=1+4=5$.
(8) 置$\{\sigma\}=\{+2,-3\}$.
(1) $(s_1,s_2)=(-6,-1)$.
(2) $I=\{1,2\}$.
(3) $J=\{1\}$.
(4) $J_1=\{1\}$, $J_2=\{\varnothing\}$; $t_1=3$, $t_2=0$.
(5) $s_1+t_1=-3$, $s_2+t_2=-1$.
(8) $\{\sigma\}=\{-2\}$.
(1) $(s_1,s_2)=(-4,-2)$.
(2) $I=\{1,2\}$.
(3) $J=\{1,3\}$.
(4) $J_1=\{1,3\}$, $J_2=\{3\}$, $t_1=9$ $t_2=1$.
(5) $s_1+t_1=5$　$s_2+t_2=-1$.
(8) 因$\{\sigma\}$中的固定变量都是零值，故步骤终止，最优解为 $x^*=(0,1,1,0,0)$, $x_0^*=5$. 计算过程如图 3.7.

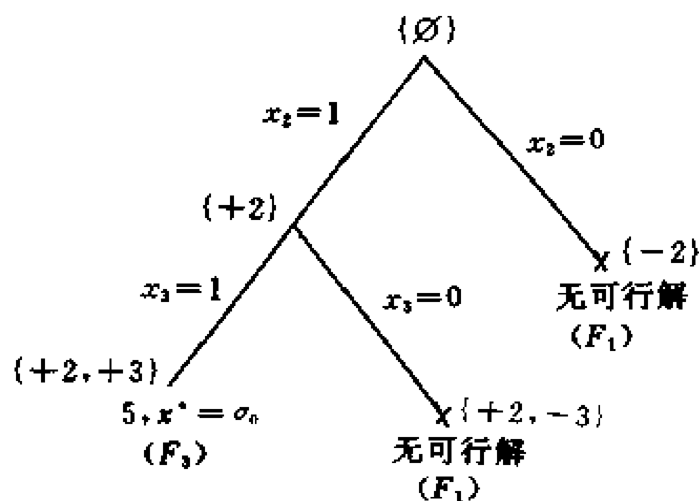



图　3.7

## 3.5 整数规划的割平面方法

### 3.5.1 参数表示式

考虑线性整数规划问题$(P)$:

$$\max \quad x_0 = \boldsymbol{cx}; \tag{3.12}$$

$$\text{s.t.} \quad \boldsymbol{Ax} = \boldsymbol{b}, \tag{3.13}$$

$$\boldsymbol{x} \geqslant \boldsymbol{0}, \tag{3.14}$$

$$\boldsymbol{x} \text{ 为整数向量}. \tag{3.15}$$

记线性规划问题(3.12)—(3.14)为$(\bar{P})$. 对$(\bar{P})$的任意基$\boldsymbol{B}$,记它的非基变量为$t_1,t_2,\cdots,t_d,d=n-m$. 以$t_j$为参数,从约束方程组(3.13)中解出基变量后,可将问题(P)写成如下的参数形式:

$$\begin{aligned}
\max \quad & x_0; \\
\text{s.t.} \quad & x_0 = a_{00} + \sum_{j=1}^{d} a_{0j}(-t_j), \\
& x_1 = a_{10} + \sum_{j=1}^{d} a_{1j}(-t_j), \\
& \ \vdots \qquad \vdots \qquad \vdots \\
& x_n = a_{n0} + \sum_{j=1}^{d} a_{nj}(-t_j), \\
& x_i \geqslant 0, \quad i = 1,2,\cdots,n, \\
& x_i \text{ 取整数值}, \quad i = 0,1,\cdots,n.
\end{aligned}$$

在上述条件中,当$x_i$是非基变量时,对应的参数表示式是恒等式$x_i=0+(-1)(-x_i)$. 记向量:

$$\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{\alpha}_j = \begin{bmatrix} a_{0j} \\ a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix} \quad (j = 0,1,\cdots,d),$$

$\boldsymbol{x}$ 为未知向量；$\boldsymbol{\alpha}_j$ 为已知向量. 则问题(P)可写成如下形式：

$$\max \quad x_0; \tag{3.16}$$

$$\text{s.t.} \quad \boldsymbol{x} = \boldsymbol{\alpha}_0 + \sum_{j=1}^{d} \boldsymbol{\alpha}_j(-t_j), \tag{3.17}$$

$$x_i \geqslant 0, \quad i = 1,2,\cdots,n, \tag{3.18}$$

$$x_i \text{ 取整数值}, i = 0,1,\cdots,n. \tag{3.19}$$

容易证明：

(1) 当 $a_{i0} \geqslant 0 \quad (i=1,2,\cdots,n)$时，则 $\boldsymbol{x}=\boldsymbol{\alpha}_0$ 是$(\tilde{\mathrm{P}})$的基本可行解.

(2) 当 $a_{0j} \geqslant 0 \quad (j=1,2,\cdots,d)$时，则对线性规划问题$(\tilde{P})$的任何可行解 $\boldsymbol{x}$，都有 $x_0 \leqslant a_{00}$. 即此时的 $a_{00}$是$(\tilde{P})$的目标函数值的一个上界.

(3) 当 $a_{i0} \geqslant 0, a_{0j} \geqslant 0 (i=1,\cdots,n, j=1,\cdots,d)$时，则 $\boldsymbol{x}=\boldsymbol{\alpha}_0$ 是$(\tilde{P})$的一个基本最优解.

(4) 当 $a_{i0} \geqslant 0, a_{0j} \geqslant 0$，且所有的 $a_{i0}$都是整数时$(i=1,\cdots,n, j=1,\cdots,d)$，则 $\boldsymbol{x}=\boldsymbol{\alpha}_0$ 是整数规划$(P)$的一个最优解.

(5) 若$(\tilde{P})$有可行解而无最优解，则$(P)$无最优解.

在参数表达式(3.17)中，假如用 $x_r$ 去替换 $t_s$，作为第 $s$ 个参变量，那么，只要从 $x_r$ 的表达式中解出 $t_s$，代入其他各式即成. 设用 $x_r$ 替换 $t_s$ 后的参数表达式为

$$\max \quad x_0;$$

$$\text{s.t.} \quad \boldsymbol{x} = \bar{\boldsymbol{a}}_0 + \sum_{j=1}^{d} \bar{\boldsymbol{a}}_j(-\bar{t}_j),$$

$$\boldsymbol{x} = (x_0, x_1, \cdots, x_n)^{\mathrm{T}};$$

$$x_i \geqslant 0 \quad (i = 1,\cdots,n),$$

$$x_i \text{ 取整数}(i = 0,1,\cdots,n).$$

其中 $x_r=\bar{t}_s, \bar{t}_j=t_j (1 \leqslant j \neq s \leqslant d)$. 则 $\bar{a}_i$ 与 $a_i$ 之间有如下的关系式：

(6) $\bar{\boldsymbol{a}}_s=-\frac{1}{a_{rs}}\boldsymbol{a}_s,\bar{\boldsymbol{a}}_j=\boldsymbol{a}_j-\frac{a_{rj}}{a_{rs}}\boldsymbol{a}_s$

$(0\leqslant j\neq s\leqslant n)$.

事实上,从

$$x_r = a_{r0} + \sum_{j\neq s} a_{rj}(-t_j) + a_{rs}(-t_s)$$

解得

$$t_s = \frac{a_{r0}}{a_{rs}} + \sum_{j\neq s}\frac{a_{rj}}{a_{rs}}(-t_j) + \frac{1}{a_{rs}}(-x_r),$$

代入 $x_i$ 的表达式中,可得

$$\begin{aligned}x_i &= \left(a_{i0}-\frac{a_{r0}}{a_{rs}}a_{is}\right) + \sum_{j\neq s}\left(a_{ij}-\frac{a_{rj}}{a_{rs}}a_{is}\right)(-t_j)\\ &\quad + \left(-\frac{a_{is}}{a_{rs}}\right)(-x_r)\\ &= \bar{a}_{i0} + \sum_{j\neq s}\bar{a}_{ij}(-\bar{t}_j) + \bar{a}_{is}(-\bar{t}_s) \quad (i=0,1,\cdots,n).\end{aligned}$$

### 3.5.2 对偶单纯形算法(dual simplex algorithm)

**定义 3.5.1** 一个非零向量 $\boldsymbol{a}=(a_0,a_1,\cdots,a_n)^{\mathrm{T}}$,若按分量的下标顺序,第一个不为零的分量是正数,则称 $\boldsymbol{a}$ 是**字典序正**,记作 $\boldsymbol{a}\overset{l}{>}\boldsymbol{0}$. 若 $-\boldsymbol{a}\overset{l}{>}\boldsymbol{0}$,则称 $\boldsymbol{a}$ 是**字典序负**,记作 $\boldsymbol{a}\overset{l}{<}\boldsymbol{0}$. 若两个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$,存在 $\boldsymbol{a}-\boldsymbol{b}\overset{l}{>}\boldsymbol{0}$,则记为 $\boldsymbol{a}\overset{l}{>}\boldsymbol{b}$. 按照字典序,所有 $n$ 维向量成为一全序集合.

计算步骤:

考虑问题(3.16)~(3.18)假设各向量 $\boldsymbol{a}_j$ 已满足条件 $\boldsymbol{a}_j\overset{l}{>}\boldsymbol{0}$($j=1,2,\cdots,d$).(如何获得这样的初始表达式,将在本节末叙述)

(1) 计算 $a_{r0}=\min\limits_{1\leqslant i\leqslant n} a_{i0}$.

若 $a_{r0}\geqslant 0$,则步骤终止. $\boldsymbol{x}=\boldsymbol{a}_0$ 便是($\tilde{\mathrm{P}}$)的最优解.

(2) 若 $a_{r0}<0$，且所有的 $a_{rj}\geqslant 0$ $(j=1,\cdots,d)$，则步骤终止，问题无可行解(因为无非负的 $x$ 满足第 $r$ 个方程)．

(3) 按字典序的大小计算：

$$\max \left\{\frac{1}{a_{rj}}\boldsymbol{a}_j \,\middle|\, a_{rj}<0, 1\leqslant j\leqslant d\right\}$$

$$=\frac{1}{a_{rs}}\boldsymbol{a}_s.$$

(4) 用 $x_r$ 替换 $t_s$ 作为第 $s$ 个参变量，得新的表达式

$$\boldsymbol{x}=\bar{\boldsymbol{a}}_0+\sum_{j=1}^{d}\bar{\boldsymbol{a}}_j(-\bar{t}_j),$$

其中 $\bar{\boldsymbol{a}}_s=-\dfrac{1}{a_{rs}}\boldsymbol{a}_s, \bar{\boldsymbol{a}}_j=\boldsymbol{a}_j-\dfrac{a_{rj}}{a_{rs}}\boldsymbol{a}_s \quad j\neq s, \bar{t}_j=t_j \quad (j\neq s), \bar{t}_s=x_r.$
用新的表达式代替原来的表达式，转(1)．

在此算法中有重要性质：

若 $\boldsymbol{a}_j\overset{l}{>}\boldsymbol{0}, j=1,\cdots,d$，则

$$\bar{\boldsymbol{a}}_j\overset{l}{>}\boldsymbol{0}, j=1,\cdots,d. \tag{3.20}$$

$$\boldsymbol{a}_0\overset{l}{>}\bar{\boldsymbol{a}}_0. \tag{3.21}$$

由(3.20)和(3.21)可知，在计算过程中，同样的表达式决不重复出现，故程序必在有限步内终止．

下面介绍一种求初始字典序正(即 $\boldsymbol{a}_j\overset{l}{>}\boldsymbol{0}$)表达式的方法，称作 $M$-方法．首先，任给一表达式 $\boldsymbol{x}=\boldsymbol{a}_0+\sum\limits_j\boldsymbol{a}_j(-t_j)$．设按字典序的大小

$$\boldsymbol{a}_s=\min\{\boldsymbol{a}_j \mid 1\leqslant j\leqslant d\}.$$

若 $\boldsymbol{a}_s\overset{l}{>}\boldsymbol{0}$，则已获字典序正的表达式．相反，在表达式的最后，我们附加一个参考条件

$$x_{n+1}=M+\sum_j(-t_j)$$

($M$ 是足够大的数). 然后,以 $x_{n+1}$替换 $t_s$ 得

$$x = \bar{a}_0 + \sum_j \bar{a}_j(-\bar{t}_j),$$

其中 $\bar{a}_s = -a_s, \bar{a}_j = a_j - a_s (1 \leqslant j \neq s \leqslant d), \bar{a}_0 = a_0 - Ma_s$. 显然,这时已有 $\bar{a}_j \overset{l}{>} 0 (j=1,\cdots,d)$. 因此就可接着应用对偶单纯形算法求线性规划问题($\bar{\mathrm{P}}$)的解.

**注意**:这时,$\bar{a}_{i0}$的形式为 $l+Mf$. 在判断 $\bar{a}_{i0} \geqslant 0$ 时,首先应该看是否满足 $f \geqslant 0$,只有当 $f=0$ 时,才看是否满足 $l \geqslant 0$.

### 3.5.3 基本割平面

考虑表达式(3.17)中的第 $i$ 个方程

$$x_i = a_{i0} + \sum_{j=1}^{d} a_{ij}(-t_j), \tag{3.22}$$

用任意的 $h(\neq 0)$乘(3.22)的两边得

$$hx_i = ha_{i0} + \sum_{j=1}^{d} ha_{ij}(-t_j), \tag{3.23}$$

则(3.23)式亦可写为

$$\lfloor h \rfloor x_i + \sum_{j=1}^{d} \lfloor ha_{ij} \rfloor t_j + (h - \lfloor h \rfloor) x_i$$

$$+ \sum_{j=1}^{d} (ha_{ij} - \lfloor ha_{ij} \rfloor) t_j = \lfloor ha_{i0} \rfloor + (ha_{i0} - \lfloor ha_{i0} \rfloor).$$

由于要求 $x_i \geqslant 0, t_j \geqslant 0$,可推得关系

$$\lfloor h \rfloor x_i + \sum_{j=1}^{d} \lfloor ha_{ij} \rfloor t_j \leqslant \lfloor ha_{i0} \rfloor + (ha_{i0} - \lfloor ha_{i0} \rfloor). \tag{3.24}$$

进一步,由于要求 $x_i$ 和 $t_j$ 是整数,因此(3.24)式的左边必须是整数. 又由于 $0 \leqslant ha_{i0} - \lfloor ha_{i0} \rfloor < 1$,则可推得关系

$$\lfloor h \rfloor x_i + \sum_{j=1}^{d} \lfloor ha_{ij} \rfloor t_j \leqslant \lfloor ha_{i0} \rfloor. \tag{3.25}$$

另一方面,若用$\lfloor h \rfloor$乘(3.22)的两边,又可得

$$\lfloor h \rfloor x_i + \sum_{j=1}^{d} \lfloor h \rfloor a_{ij} t_j = \lfloor h \rfloor a_{i0}. \qquad (3.26)$$

最后，由(3.25)—(3.26)可导得关系式

$$\sum_{j=1}^{d} (\lfloor h \rfloor a_{ij} - \lfloor h a_{ij} \rfloor) t_j \geqslant \lfloor h \rfloor a_{i0} - \lfloor h a_{i0} \rfloor. \qquad (3.27)$$

称条件(3.27)为基本割平面. 方程(3.22)为导出此割平面的诱导方程. 它所对应的行 $i$ 称为诱导行.

对非负整数解来说，割平面条件(3.27)是条件(3.22)的推论. 即凡是满足(3.22)的非负整数解自然也满足(3.27). 然而满足(3.22)的非负解就不一定能满足(3.27). 例如，当 $a_{i0}>0$，但不是整数时(3.27)的右边，一般都为正数. 则满足(3.22)的非负解 $x_i=a_{i0}, t_j=0(j=1,\cdots,d)$ 就不满足(3.27). 割平面的作用，就是能从定义域(3.17)～(3.18)中割去一部分非负解，但不割去非负整数解. 根据 $h$ 的不同取值，下面将导出各种形式的割平面.

### 3.5.4 分数割平面算法(fractional cutting plane algorithm)

在基本割平面(3.27)中，让 $h=1$，则可得

$$\sum_{j=1}^{d} (a_{ij} - \lfloor a_{ij} \rfloor) t_j \geqslant a_{i0} - \lfloor a_{i0} \rfloor.$$

记 $r_{ij}=a_{ij}-\lfloor a_{ij} \rfloor (j=0,1,\cdots,d)$，则上式可写为

$$\sum_{j=1}^{d} r_{ij} t_j \geqslant r_{i0}.$$

引进非负的松弛变量 $s_i$，得条件

$$s_i = -r_{i0} + \sum_{j=1}^{d} r_{ij} t_j, \qquad (3.28)$$

称(3.28)为分数割平面. 方程(3.28)中的各个系数正好是(3.22)中各个系数的分数部分. 因为

$$x_i = \left( \lfloor a_{i0} \rfloor - \sum_{j=1}^{d} \lfloor a_{ij} \rfloor t_j \right) - \left( -r_{i0} + \sum_{j=1}^{d} r_{ij} t_j \right)$$

$$=\left(\lfloor a_{i0}\rfloor-\sum_{j=1}^{d}\lfloor a_{ij}\rfloor t_j\right)-s_i.$$

所以当 $x_i, t_j$ 都取非负整数值时，$s_i$ 自然也取非负整数值.

求解整数规划($P$)的计算步骤：

(1) 利用对偶单纯形方法，求松弛问题($\tilde{P}$)的最优解. 若($\tilde{P}$)没有最优解，则步骤终止，($P$)也没有最优解. 若($\tilde{P}$)有最优解，设所求得的达到最优时的表达式为

$$\boldsymbol{x}=\boldsymbol{a}_0+\sum_{j=1}^{d}\boldsymbol{a}_j(-t_j)$$

(即此时有 $a_{i0}\geqslant 0, a_{0j}\geqslant 0, i=1,\cdots,n, j=1,\cdots,d$).

(2) 若所有的 $a_{i0}$ 都是整数，则步骤终止，$\boldsymbol{x}=\boldsymbol{a}_0$ 便是($P$)的最优解. 相反，设 $a_{l0}$ 是最头上的非整数，即

$$l=\min\{i\,|\,a_{i0}\text{ 不是整数}, i=0,1,\cdots,n\}.$$

取诱导方程为

$$x_l=a_{l0}+\sum_{j=1}^{d}a_{lj}(-t_j)\quad(\text{即取诱导行为 } l).$$

(3) 从诱导方程导出分数割平面

$$s_l=-r_{l0}-\sum_{j=1}^{d}r_{lj}(-t_j),$$

其中　$r_{lj}=a_{lj}-\lfloor a_{lj}\rfloor$.

(4) 按字典序的大小计算

$$\max\left\{\frac{1}{-r_{lj}}\boldsymbol{a}_j\,\middle|\,r_{lj}>0, 1\leqslant j\leqslant d\right\}=\frac{1}{-r_{ls}}\boldsymbol{a}_s$$

(5) 用 $s_l$ 替换 $t_s$，得新的表达式

$$\boldsymbol{x}=\bar{\boldsymbol{a}}_0+\sum_{j=1}^{d}\bar{\boldsymbol{a}}_j(-\bar{t}_j),$$

其中　$\bar{t}_j=t_j$，当 $j\neq s$ 时，而 $\bar{t}_s=s_l$，

$$\bar{\boldsymbol{a}}_s=\frac{1}{r_{ls}}\boldsymbol{a}_s,\ \bar{\boldsymbol{a}}_j=\boldsymbol{a}_j-\frac{r_{lj}}{r_{ls}}\boldsymbol{a}_s\quad(j\neq s).$$

(6) 若 $\bar{a}_{i0}\geqslant 0(i=1,\cdots,n)$，则转(2). 相反，转(1)，求改进后的(即加上割平面条件后的)松弛问题的最优解.

**注意**，引进割平面条件，只是为了作一次参数变换，实际上并不需要把它加到约束方程组(3.17)中，因为我们已经知道它是诱导方程的推论.

**例 3.5.2** 求解

$$
\begin{aligned}
\min\quad & 4x+5y;\\
\text{s.t.}\quad & 3x+2y\geqslant 7,\\
& x+4y\geqslant 5,\\
& 3x+y\geqslant 2,\\
& x,y\ \text{为非负整数}.
\end{aligned}
$$

**解** 写成参数形式为：

$$
\begin{aligned}
\max\quad & x_0;\\
\text{s.t.}\quad & x_0=-4x_4-5x_5,\\
& x_1=-7+3x_4+2x_5,\\
& x_2=-5+x_4+4x_5,\\
& x_3=-2+3x_4+x_5,\\
& x_1,x_2,\cdots,x_5\ \text{为非负整数},x_0\ \text{为整数}.
\end{aligned}
$$

列成如下单纯形表：

$$
\begin{array}{c}
\begin{array}{c}
\\ x_0\\ x_1\\ x_2\\ x_3\\ x_4\\ x_5
\end{array}
\begin{array}{ccc}
1 & -x_4 & -x_5\\
\hline
\end{array}
\end{array}
$$

$$
\begin{array}{c|ccc}
 & 1 & -x_4 & -x_5\\
\hline
x_0 & 0 & 4 & 5\\
x_1 & -7 & -3 & -2\\
x_2 & -5 & -1 & -4\\
x_3 & -2 & -3 & -1\\
x_4 & & -1 & \\
x_5 & & & -1\\
\hline
 & \boldsymbol{a}_0 & \boldsymbol{a}_1 & \boldsymbol{a}_2
\end{array},
$$

现在已满足 $a_j \overset{l}{>} 0(j\geqslant 1)$的要求.

(1) 用对偶单纯形方法求线性规划最优解.

| | 1 | $-x_4$ ↓ | $-x_5$ | 1 | $-x_1$ | $-x_5$ ↓ | 1 | $-x_1$ | $-x_2$ |
|---|---|---|---|---|---|---|---|---|---|
| $x_0$ | 0 | 4 | 5 | $-\frac{28}{3}$ | $\frac{4}{3}$ | $\frac{7}{3}$ | $-\frac{112}{10}$ | $\frac{11}{10}$ | $\frac{7}{10}$ |
| → $x_1$ | $-7$ | $-3^*$ | $-2$ | 0 | $-1$ | 0 | 0 | $-1$ | 0 |
| $x_2$ | $-5$ | $-1$ | $-4$ | → $-\frac{8}{3}$ | $-\frac{1}{3}$ | $-\frac{10^*}{3}$ | 0 | 0 | $-1$ |
| $x_3$ | $-2$ | $-3$ | $-1$ | 5 | $-1$ | 1 | $\frac{42}{10}$ | $-\frac{11}{10}$ | $\frac{3}{10}$ |
| $x_4$ | 0 | $-1$ | 0 | $\frac{7}{3}$ | $-\frac{1}{3}$ | $\frac{2}{3}$ | $\frac{18}{10}$ | $-\frac{4}{10}$ | $\frac{2}{10}$ |
| $x_5$ | 0 | 0 | $-1$ | 0 | 0 | $-1$ | $\frac{8}{10}$ | $\frac{1}{10}$ | $-\frac{3}{10}$ |
| | (Ⅰ) | | | (Ⅱ) | | | (Ⅲ) | | |

(2) 因 $a_{00}$不是整数,选诱导方程为

$$x_0=-\frac{112}{10}-\frac{11}{10}x_1-\frac{7}{10}x_2.$$

(3) 导出割平面 $s_0=-\frac{8}{10}-\frac{1}{10}(-x_1)-\frac{7}{10}(-x_4)$.

(4) 求 $\max\left\{\frac{1}{-\frac{1}{10}}a_1,\frac{1}{-\frac{7}{10}}a_2\right\}=-\frac{10}{7}a_2$.

(5) 以 $s_0$ 替换 $x_2$ 得表(Ⅳ). 重复步骤(2)到步骤(5),得表(Ⅴ)和表(Ⅵ).(表中附"#"的是诱导行,附"↓"是被替换的参变量).

| | | | ↓ | | ↓ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | $-x_1$ | $-s_0$ | 1 | $-x_1$ | $-s_2$ | 1 | $-s_0'$ | $-s_2$ |
| $x_0$ | $-12$ | 1 | 1 | # $-12\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{7}{4}$ | $-13$ | 1 | 1 |
| $x_1$ | | $-1$ | | | $-1$ | | 1 | $-\frac{4}{3}$ | 1 |
| $x_2$ | $\frac{8}{7}$ | $\frac{1}{7}$ | $-\frac{10}{7}$ | $\frac{6}{4}$ | $\frac{2}{4}$ | $-\frac{10}{4}$ | 1 | $\frac{2}{3}$ | $-3$ |
| $x_3$ | $\frac{27}{7}$ | $-\frac{8}{7}$ | $\frac{3}{7}$ | $\frac{15}{4}$ | $-\frac{5}{4}$ | $\frac{3}{4}$ | 5 | $-\frac{5}{3}$ | 2 |
| $x_4$ | $\frac{11}{7}$ | $-\frac{3}{7}$ | $\frac{2}{7}$ | $\frac{6}{4}$ | $-\frac{2}{4}$ | $\frac{2}{4}$ | 2 | $-\frac{2}{3}$ | 1 |
| $x_5$ | $\frac{8}{7}$ | $\frac{1}{7}$ | $-\frac{3}{7}$ | $\frac{5}{4}$ | $\frac{1}{4}$ | $-\frac{3}{4}$ | 1 | $\frac{1}{3}$ | $-1$ |
| | | (Ⅳ) | | | (Ⅴ) | | | (Ⅵ) | |

最终求得问题的最优解为 $x=x_4=2, y=x_5=1$，最优目标函数的值为 13，表(Ⅰ)至(Ⅵ)的变换过程如图 3.8 中点①至点⑥.

第一个割平面为

$$s_0=-\frac{8}{10}+\frac{1}{10}x_1+\frac{7}{10}x_2=-5+x_4+3x_5\geqslant 0,$$

加上它后，使点③→点④.

第二个割平面为

$$s_2=-\frac{1}{7}+\frac{1}{7}x_1+\frac{4}{7}s_0=-4+x_4+2x_5\geqslant 0,$$

加上它后，使点④→点⑤.

第三个割平面为

$$s_0'=-\frac{3}{4}+\frac{3}{4}x_1+\frac{3}{4}s_2=-9+3x_4+3x_5\geqslant 0,$$

加上它后，使点⑤→点⑥.

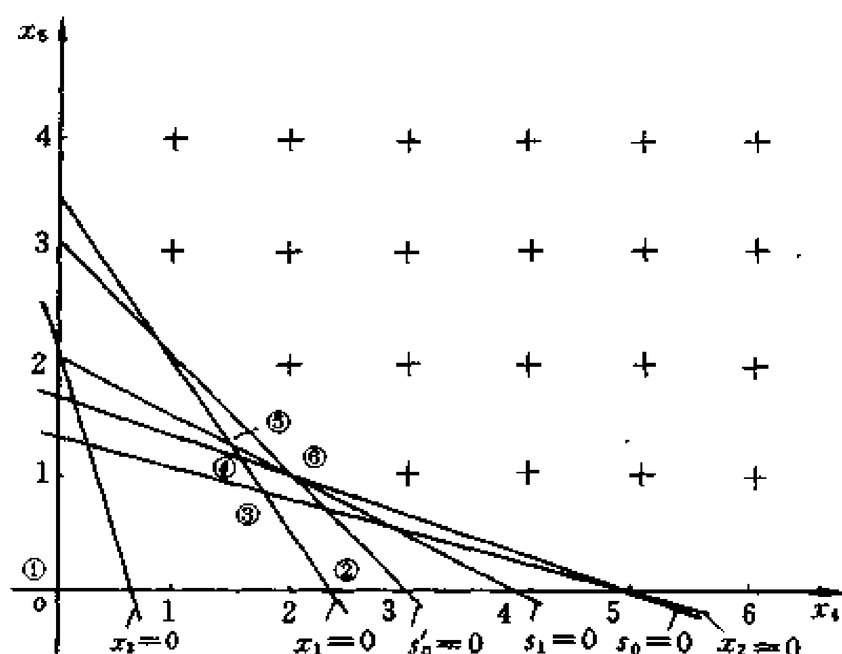



图 3.8

## 3.6 分解算法(decomposition algorithm)

设线性混合整数规划问题($P$):

$$\max \quad x_0 = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{d}^{\mathrm{T}}\boldsymbol{y}; \tag{3.29}$$

$$\text{s.t.} \quad \boldsymbol{Ax} + \boldsymbol{By} = \boldsymbol{b}, \tag{3.30}$$

$$\boldsymbol{x} \geqslant \boldsymbol{0}, \quad \boldsymbol{k} \geqslant \boldsymbol{y} \geqslant \boldsymbol{0}, \tag{3.31}$$

$$\boldsymbol{y}\ \text{是整数向量}. \tag{3.32}$$

其中 $x_0$ 是变量,$\boldsymbol{c}^{\mathrm{T}}=(c_1,c_2,\cdots,c_p)$,$\boldsymbol{d}^{\mathrm{T}}=(d_1,d_2,\cdots,d_q)$,$\boldsymbol{x}=(x_1,x_2,\cdots,x_p)^{\mathrm{T}}$,$\boldsymbol{y}=(y_1,y_2,\cdots,y_q)^{\mathrm{T}}$,$\boldsymbol{A}=(a_{ij})$是 $m\times p$ 矩阵,$\boldsymbol{B}=(b_{ij})$是 $m\times q$ 矩阵,$\boldsymbol{b}=(b_1,b_2,\cdots,b_m)^{\mathrm{T}}$,$\boldsymbol{k}=(k_1,k_2,\cdots,k_q)^{\mathrm{T}}$. 令

$$S=\{\boldsymbol{y}\,|\,\boldsymbol{k}\geqslant\boldsymbol{y}\geqslant\boldsymbol{0},\boldsymbol{y}\ \text{是整数向量}\},$$

$$F=\{\boldsymbol{y}\,|\,\boldsymbol{y}\in S,\text{且使}:\boldsymbol{Ax}=\boldsymbol{b}-\boldsymbol{By},\boldsymbol{x}\geqslant\boldsymbol{0}\ \text{有解}\}.$$

对 $\boldsymbol{y}\in F$,定义线性规划问题 $P(\boldsymbol{y})$如下:

求 $x_0(y)=\max\ x_0$；　　(3.33)

$$\text{s.t.}\quad \left.\begin{aligned} &x_0=\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{d}^{T}\boldsymbol{y},\\ &\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}-\boldsymbol{B}\boldsymbol{y},\\ &\boldsymbol{x}\geqslant 0. \end{aligned}\right\} \tag{3.34}$$

则问题$(P)$可写为：

$$\begin{aligned} &\max\quad x_0(\boldsymbol{y});\\ &\text{s.t.}\quad \boldsymbol{y}\in F, \end{aligned}$$

根据线性规划对偶理论，$P(\boldsymbol{y})$的对偶规划为$D(\boldsymbol{y})$：

$$\min\quad \boldsymbol{d}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{b}; \tag{3.35}$$

$$\text{s.t.}\quad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\geqslant \boldsymbol{c}^{\mathrm{T}}, \tag{3.36}$$

其中$\boldsymbol{u}^{\mathrm{T}}=(u_1,u_2,\cdots,u_m)$. 令

$$U=\{\boldsymbol{u}\,|\,\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\geqslant \boldsymbol{c}^{\mathrm{T}}\}.$$

若$U=\varnothing$，则对任何的$\boldsymbol{y}\in F$，必有$x_0(\boldsymbol{y})=+\infty$. 因此，或者$F=\varnothing$，或者$(P)$无最优解；若$U\neq\varnothing$，这时，若$P(\boldsymbol{y})$又有可行解，则$P(\boldsymbol{y})$和$D(\boldsymbol{y})$都有最优解，且有：

$$\begin{aligned} &x_0(\boldsymbol{y})=\min\ \boldsymbol{d}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{u}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}+\boldsymbol{u}^{\mathrm{T}}\boldsymbol{b};\\ &\text{s.t.}\ \boldsymbol{u}\in U. \end{aligned}$$

若$P(\boldsymbol{y})$无可行解，则$D(\boldsymbol{y})$无最优解.

根据多面集表示定理 2.1.13，对任意的$\boldsymbol{u}\in U$，可写为：

$$\boldsymbol{u}=\sum_i\lambda_i\boldsymbol{u}^i+\sum_j\mu_j\boldsymbol{v}^j,\quad \lambda_i\geqslant 0,\mu_j\geqslant 0,\sum_i\lambda_i=1.$$

其中$\boldsymbol{u}^i$为$U$的极点(基本可行解)，$\boldsymbol{v}^j$为$U$的极方向.

根据对偶理论，若$U\neq\varnothing$，则$\boldsymbol{y}\in F$的充要条件是$D(\boldsymbol{y})$有最优解. 而$D(\boldsymbol{y})$有最优解的充要条件是：对所有的$\boldsymbol{v}^j$，满足

$$-(\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}+(\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{b}\geqslant 0,\qquad (\boldsymbol{y}\in S)$$

$$x_0(\boldsymbol{y})=\min_j\ \boldsymbol{d}^{\mathrm{T}}\boldsymbol{y}-(\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}+(\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{b}.$$

因此，问题$(P)$可归结为如下问题$(I.P.)$：

$$\max \quad x_0; \tag{3.37}$$

$$\left.\begin{aligned} \text{s.t.} \quad & -(\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{By} + (\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{b} \geqslant 0, \text{对所有的 } \boldsymbol{v}^j, \\ & \boldsymbol{d}^{\mathrm{T}}\boldsymbol{y} - (\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{By} + (\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{b} \geqslant x_0, \text{对所有的 } \boldsymbol{u}^j, \\ & \boldsymbol{0} \leqslant \boldsymbol{y} \leqslant \boldsymbol{k}, \quad (\text{即 } \boldsymbol{y} \in S). \end{aligned}\right\} \tag{3.38}$$

若问题$(I.P.)$无可行解，则对任何的 $\boldsymbol{y}\in S$，$D(\boldsymbol{y})$无最优解，因此，$P(\boldsymbol{y})$无可行解，即$(P)$无可行解；若$(I.P.)$有可行解，则必有最优解，记它的最大值为 $x_0(I.P.)$，设求得的最优解为 $\boldsymbol{y}^0\in F$. 则只要再解一次如下的线性规划问题：

$$\begin{aligned} \max \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}; \\ \text{s.t.} \quad & \boldsymbol{Ax} = \boldsymbol{b} - \boldsymbol{By}^0,\ \boldsymbol{x} \geqslant \boldsymbol{0}. \end{aligned}$$

便可求得混合整数规划$(P)$的最优解.

以上说明了分解算法的原理，但是，问题$(I.P.)$实际上是不知道的. 下面介绍的是应用松弛技术的计算过程.

设 $Q$ 为极点 $\boldsymbol{u}^j$ 的任意一个子集，$R$ 为极方向 $\boldsymbol{v}^j$ 的任意一个子集，定义松弛问题 $P(Q,R)$如下：

$$\max \quad x_0; \tag{3.39}$$

$$\left.\begin{aligned} \text{s.t.} \quad & -(\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{By} + (\boldsymbol{v}^j)^{\mathrm{T}}\boldsymbol{b} \geqslant 0, \text{ 对所有的 } \boldsymbol{v}^j \in R, \\ & (\boldsymbol{d})^{\mathrm{T}}\boldsymbol{y} - (\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{By} + (\boldsymbol{u}^j)^{\mathrm{T}}\boldsymbol{b} \geqslant x_0, \text{对所有的 } \boldsymbol{u}^j \in Q, \\ & \boldsymbol{0} \leqslant \boldsymbol{y} \leqslant \boldsymbol{k}. \end{aligned}\right\} \tag{3.40}$$

若问题 $P(Q,R)$无可行解，则$(P)$也无可行解；若 $P(Q,R)$有可行解，则必有最优解，记其最大值为 $x_0(Q,R)$. 显然有：$x_0(Q,R)\geqslant x_0(I,P)$. 对任一 $\boldsymbol{y}\in S$，若 $U\neq\varnothing$，则或者 $P(\boldsymbol{y})$无可行解，或者 $P(\boldsymbol{y})$有最优解，最大值为 $x_0(\boldsymbol{y})$. 且对任意的 $Q$ 和 $R$ 满足关系：

$$x_0(\boldsymbol{y}) \leqslant x_0(I,P) \leqslant x_0(Q,R).$$

因此，若有某一个 $\boldsymbol{y}^0\in S$，且有某子集 $Q,R$. 使得 $x_0(\boldsymbol{y}^0)=x_0(Q,R)$，则 $\boldsymbol{y}^0$ 便是问题$(I.P.)$的最优解. 这就是下述算法的主要依据.

计算步骤：

(1) 任取一 $y\in S$，用对偶单纯形法，求解互为对偶的线性规划问题 $P(y)$和 $D(y)$.

(i) 若 $D(y)$无可行解，则步骤终止，问题$(P)$无最优解.

(ii) 若 $D(y)$有可行解，但无最优解，则必可找到 $U$ 的一个极点 $u^i$，及一个极方向 $v^j$，使得：$-(v^j)^T By+(v^j)^T b<0$. 以此 $u^i$ 和 $v^j$ 构成初始的子集 $Q$ 和 $R$.

(iii) 若 $D(y)$有最优解，设求得 $D(y)$的一个基本最优解(极点)为 $u^i$，求得 $P(y)$的基本最优解为 $x^*$. 则以此 $u^i$ 构成初始的子集 $Q$，$R:=\varnothing$.

(2) 解整数规划 $P(Q,R)$.

(i) 若 $P(Q,R)$无可行解，则步骤终止，$(P)$无可行解.

(ii) 若 $P(Q,R)$有最优解，设为 $y^*$，最大值为 $x_0(Q,R)$.

(3) 解线性规划 $D(y^*)$和 $P(y^*)$.

(i) 若 $D(y^*)$无最优解(即 $P(y^*)$无可行解)，则可找到 $U$ 的一个极点 $u^i$，及一个极方向 $v^j$，使 $-(v^j)^T By^*+(v^j)^T b<0$. 将 $v^j$ 加进子集 $R$. 若 $d^T y^*-(u^i)^T By^*+(u^i)^T b<x_0(Q,R)$，则也将 $u^i$ 加进子集 $Q$. 然后转(1).

(ii) 若 $D(y^*)$有最优解，设为 $u^i$. 设 $P(y^*)$的最优解为 $x^*$. 设 $P(y^*)$的最大值为 $x_0(y^*)$. 这时，若 $x_0(y^*)=x_0(Q,R)$，则步骤终止，$(x_0(y^*),x^*,y^*)$便是$(P)$的最优解；若 $x_0(y^*)<x_0(Q,R)$，则将此 $u^i$ 加进子集 $Q$，然后转(2).

每执行一次(3)，子集 $Q$ 或 $R$ 中，将增加一个元素 $u^i$ 或 $v^j$，它们满足条件：

$$d^T y^* - (u^i)^T By^* + (u^i)^T b < x_0(Q,R).$$

或

$$(-v^j)^T By^* + (v^j)^T b < 0.$$

由此可推知：每次增加的 $u^i$ 或 $v^j$，都不是原来的 $Q$ 或 $R$ 中的极点

或极方向.因此,过程必在有限步内终止.

## 3.7 集合覆盖和分解问题的解法

三个最基本的 0—1 规划问题是:

覆盖问题(covering problem)

$$\min \quad \boldsymbol{cx}; \tag{3.41}$$

$$\left.\begin{aligned} \text{s.t.} \quad & \boldsymbol{Ax} \geqslant \boldsymbol{1}, \\ & \boldsymbol{x} \text{ 为 } 0\text{—}1 \text{ 向量}. \end{aligned}\right\} \tag{3.42}$$

分解问题(partition problem)

$$\min \quad \boldsymbol{cx}; \tag{3.43}$$

$$\left.\begin{aligned} \text{s.t.} \quad & \boldsymbol{Ax} = \boldsymbol{1}, \\ & \boldsymbol{x} \text{ 为 } 0\text{—}1 \text{ 向量}. \end{aligned}\right\} \tag{3.44}$$

无关子族问题(packing problem)

$$\max \quad \boldsymbol{cx}; \tag{3.45}$$

$$\left.\begin{aligned} \text{s.t.} \quad & \boldsymbol{Ax} \leqslant \boldsymbol{1}, \\ & \boldsymbol{x} \text{ 为 } 0\text{—}1 \text{ 向量}. \end{aligned}\right\} \tag{3.46}$$

其中 $\boldsymbol{A}$ 为 0—1 矩阵,$\boldsymbol{c}$ 是分量都为正的行向量.$\boldsymbol{1}$ 表示分量都为 1 的向量.图论和网络理论中的很多极值问题,都可归结成上述 0—1 规划问题.求解这些问题的难易程度和矩阵 $\boldsymbol{A}$ 的结构形式有关.根据特殊的结构形式,建立特殊的有效算法,就是组合最优化所研究的主要内容.它把组合论、图论、线性规划密切地联系在一起.这里,只介绍覆盖问题的一般解法.首先说明,分解问题可以化成如下的覆盖问题:

$$\min \quad \sum_{j=1}^{n}(c_j + Lt_j)x_j; \tag{3.47}$$

$$\left.\begin{aligned} \text{s.t.} \quad & \sum_{j=1}^{n} a_{ij}x_j \geqslant 1 \quad (i = 1,\cdots,m), \\ & x_j \text{ 取 } 0 \text{ 或 } 1. \end{aligned}\right\} \tag{3.48}$$

其中 $$t_j = \sum_{i=1}^{m} a_{ij}, L = \sum_{j=1}^{n} c_j + 1.$$

事实上，设 $x^*$ 是分解问题的任意可行解，$\hat{x}$ 是覆盖问题而不是分解问题的可行解，即 $Ax^* = 1, A\hat{x} \geqslant 1$，且

$$\hat{I} = \{i \mid \sum_{j=1}^{n} a_{ij} \hat{x}_j > 1\} \neq \varnothing.$$

则 $$\sum_{j=1}^{n} (c_j + Lt_j) x_j^* = \sum_{j=1}^{n} c_j x_j^* + Lm < L(m+1).$$

但是

$$\sum_{j=1}^{n} (c_j + Lt_j) \hat{x}_j \geqslant \sum_{j=1}^{n} c_j \hat{x}_j + L(m + |\hat{I}|) \geqslant L(m+1).$$

因此，只要分解问题有可行解，上述覆盖问题的最优解必在分解问题的可行解中达到.

**覆盖问题的化简规则** 对覆盖问题，常常可根据以下的规则，预先确定某些 $x_j$ 的值，或者去掉某些多余的约束条件. 这些规则称为化简规则.

规则 1 若 $A$ 的某一行 $a_i$ 是一个单位向量，即 $a_{ik}=1, a_{ij}=0$，对所有的 $j \neq k$. 则 $x_k$ 必须取 1. 由于 $x_k=1$，则在 $A$ 的第 $k$ 列中，凡是元素为 1 的行，条件都已得到满足. 因此，可将这些行和第 $k$ 列去掉.

规则 2 若 $A$ 中的行 $a_t$ 和 $a_p$，使得 $a_t \geqslant a_p$，则第 $t$ 个条件可以去掉. 因为它是第 $p$ 个条件的自然结果.

规则 3 若有 $A$ 的某些列指标集 $S$，以及某一列指标 $k \notin S$ 使得：

$$\sum_{j \in S} a_{ij} \geqslant a_{ik} \quad (i = 1, \cdots, m),$$

$$\sum_{j \in S} c_j \leqslant c_k,$$

则显然第 $k$ 列可以去掉，即 $x_k$ 可取为 0.

**例 3.7.1** 设覆盖问题如下：

| 行 \ $c_j$ | 10 | 5 | 8 | 6 | 9 | 13 | 11 | 4 | 6 |
|---|---|---|---|---|---|---|---|---|---|
| $a_1$ | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| $a_2$ | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| $a_3$ | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| $a_4$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| $a_5$ | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| $a_6$ | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| $a_7$ | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| 列 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

（第一行为 $c=$，$a_1$ 至 $a_7$ 行为 $A=$）

由规则 2，因 $\boldsymbol{a}_7 \geqslant \boldsymbol{a}_4$，可去掉第 7 行.

由规则 1，因 $\boldsymbol{a}_4$ 是单位向量，可取 $x_4=1$，并去掉第 4 行和第 4 列.

由规则 2，因 $\boldsymbol{a}_1 \geqslant \boldsymbol{a}_2$ 可去掉 $\boldsymbol{a}_1$.

由规则 3，因 $a_{i2} \geqslant a_{i7}(i \neq 1,4,7)$，且 $c_2 < c_7$，故可去掉第 7 列.

由规则 3，因 $a_{i3} \geqslant a_{i1}(i \neq 1,4,7)$，且 $c_3 < c_1$，故可去掉第 1 列.

由规则 3，因 $a_{i5} \geqslant a_{i6}(i \neq 1,4,7)$，且 $c_5 < c_6$，故可去掉第 6 列.

由规则 3，因 $a_{i9} \geqslant a_{i5}(i \neq 1,4,7)$，且 $c_9 < c_5$，故可去掉第 5 列.

由规则 2，由于已去掉 1,4,5,6,7 列，这时 $a_6 \geqslant a_5$，可去掉 $a_6$.

由规则 3，因 $a_{i8} \geqslant a_{i2}(i \neq 1,4,6,7)$，且 $c_8 < c_2$，故可去掉第 2 列.

最后，化简到如下覆盖问题：

| $c=$ |  | 8 | 4 | 6 |
|---|---|---|---|---|

$$A=\begin{array}{|c|c|c|c|}\hline a_2 & 1 & 1 & 0 \\ \hline a_3 & 0 & 1 & 1 \\ \hline a_5 & 1 & 0 & 1 \\ \hline \end{array}$$

| | 3 | 8 | 9 |
|---|---|---|---|

此问题的最优解为 $x_4=x_8=x_9=1$，目标函数值为 $6+4+6=16$.

**定义 3.7.2 基本覆盖**

对 $x$ 的分量的指标任一子集合 $J$，定义向量 $\boldsymbol{x}(J)$ 如下：

$$\boldsymbol{x}(J)=(x_1,\cdots,x_n)^{\mathrm{T}},$$

$$x_j=\begin{cases}1, & j\in J,\\ 0, & j\notin J.\end{cases}$$

若 $\boldsymbol{x}(J)$ 是覆盖问题的一个可行解，则称 $J$ 为一个覆盖，假如在覆盖 $J$ 中，有一个指标 $j^*$，使得 $J-\{j^*\}$ 仍是一个覆盖，即满足：

$$\sum_{j\in J}a_{ij}-a_{ij^*}\geqslant 1,(i=1,\cdots,m),$$

则称 $j^*$ 是过剩指标. 一个不含过剩指标的覆盖，称为**基本覆盖**；否则称为**过剩覆盖**. 覆盖 $J$ 中的一个指标 $j'$，它不是过剩的充要条件为：

$$I(j')=\left\{i\,\Big|\sum_{j\in J}a_{ij}-a_{ij'}=0\right\}\neq\varnothing.$$

因为 $\boldsymbol{c}>\boldsymbol{0}$，故覆盖问题的最优解必对应于一个基本覆盖.

记覆盖问题为$(P)$，它的松弛问题为$(\tilde{P})$：

$$\min\quad \boldsymbol{cx}; \tag{3.49}$$

$$\left.\begin{aligned}\text{s.t.}\quad & \boldsymbol{Ax}\geqslant \boldsymbol{1},\\ & \boldsymbol{x}\geqslant \boldsymbol{0}.\end{aligned}\right\} \tag{3.50}$$

对$(\tilde{P})$的任一可行解 $\tilde{\boldsymbol{x}}$，定义向量 $\bar{\boldsymbol{x}}$，它的分量是：

$$\bar{x}_j=\min\{1,\tilde{x}_j\}\quad(j=1,\cdots,n).$$

由于 $\boldsymbol{A}$ 是一个 0—1 矩阵，$\bar{\boldsymbol{x}}$ 也是$(\tilde{P})$的可行解. 设 $\tilde{\boldsymbol{x}}^*$ 是$(\tilde{P})$的一

个基本最优解. 则因为 $c>0$,故必有:

$$0 \leqslant \tilde{x}_j^* \leqslant 1 \quad (j=1,\cdots,n).$$

定义向量 $\bar{x}^*$ 的分量如下:

$$\bar{x}_j^* = \lceil \tilde{x}_j^* \rceil \quad (j=1,\cdots,n),$$

其中$\lceil \tilde{x}_j^* \rceil$表示不小于 $\tilde{x}_j^*$ 的最小整数,显然,$\bar{x}^*$ 是覆盖问题$(P)$的一个可行解. 记 $\bar{J}^* = \{j \mid \bar{x}_j^* = 1\}$,则 $\bar{J}^*$ 是一个覆盖,但不一定是基本覆盖. 从 $\bar{J}^*$ 中,逐个地减去过剩指标后,必可得到一个基本覆盖 $J^*$. 当然,用不同方式减过剩指标,可能获得不同的基本覆盖.

**性质 1** 若 $J$ 是一个基本覆盖,则 $\boldsymbol{x}(J)$是$(\tilde{P})$的一个基本可行解.

我们进一步讨论对应于基本可行解 $\boldsymbol{x}(J)$的基矩阵形式,因为 $J$ 是一个基本覆盖,则经过行和列的适当排列后,矩阵 $\boldsymbol{A}$ 和向量 $\boldsymbol{c}$ 必可表成如下的形式:

$$\boldsymbol{A} = \left.\begin{array}{c} \overbrace{\quad\quad\quad}^{J\text{ 中的列}} \quad \overbrace{\quad\quad\quad}^{J\text{ 以外的列}} \\ \left[\begin{array}{ccc:c} 1 & & & \\ & \ddots & & \boldsymbol{A}_{12} \\ & & 1 & \\ \hdashline & \boldsymbol{A}_{21} & & \boldsymbol{A}_{22} \end{array}\right] \end{array}\right\},$$

$$\boldsymbol{c} = (\quad \boldsymbol{c}_J \quad\quad \boldsymbol{c}_N \quad).$$

线性规划$(\tilde{P})$化成标准形式为:

$$\begin{aligned} \min \quad & \boldsymbol{cx}; \\ \text{s.t.} \quad & \boldsymbol{Ax} - \boldsymbol{Iy} = \boldsymbol{1}, \\ & \boldsymbol{x} \geqslant \boldsymbol{0}, \\ & \boldsymbol{y} \geqslant \boldsymbol{0}. \end{aligned}$$

因此,对应于 $x(J)$的基矩阵可取成如下的下三角形式:

$$
B(J)=\left[\begin{array}{cccc:cccc}
1 & & & & & & & \\
& 1 & & & & & O & \\
& & \ddots & & & & & \\
& & & 1 & & & & \\
\hdashline
& & & & -1 & & & \\
& & A_{21} & & & -1 & & \\
& & & & & & \ddots & \\
& & & & & & & -1
\end{array}\right]
$$

（上方标注：左侧列为“$J$ 中的列”，右侧列为“$-I$ 中的列”）

**性质 2** 上述形式的基矩阵 $B(J)$，满足 $B(J)^{-1}=B(J)$.

根据单纯形法，若基矩阵 $B(J)$满足条件：

$$c_B B^{-1}\bar{A}\leqslant\bar{c},$$

其中 $c_B=(c_J,0)$，$B^{-1}=B(J)^{-1}=B(J)$，$\bar{A}=(A,-I)$，$\bar{c}=(c_J,c_N,0)$. 则 $x(J)$便是$(\bar{P})$的基本最优解，因此，也是覆盖问题$(P)$的一个最优解. 用 $N$ 表示子矩阵 $A_{12}$所取的列指标集合，用 $P_j(j\in N)$表示 $A_{12}$中的列向量，则上述有关最优性的充分条件可写为：
$c_J P_j-c_j\leqslant 0$，对所有的 $j\in N$.

若设 $x(J)$不是$(\bar{P})$的最优解，记 $Q=\{j\,|\,c_J P_j-c_j>0,j\in N\}$. 这时，若有另一个基本覆盖 $J^*$，使得

$$cx(J^*)<cx(J)\text{，则 } x(J^*)=(x_1^*,\cdots,x_n^*)^{\mathrm{T}}$$

必须满足：

$$\sum_{j\in Q}x_j^*\geqslant 1.$$

称条件：$\sum_{j\in Q}x_j\geqslant 1$. 为关于基本覆盖 $J$ 的一个割平面.

覆盖问题割平面算法步骤.

据前所述，对覆盖问题$(P)$：
$\min\{x_0\,|\,x_0=cx,Ax\geqslant 1,x$ 是 0—1 向量$\}$.
有如下的步骤：

(1) 若 $A$ 中有一行全为 0，则步骤终止，问题无可行解. 否则，

任给一覆盖，作为初始的记录 $J$，置 $x_0^*=cx(J)$.

(2) 利用覆盖问题的化简规则 1，2，3，化简问题($P$)和矩阵 $\boldsymbol{A}$. 假如经过化简后，已完全确定问题($P$)的最优解，则步骤终止. 否则，进行(3).

(3) 求松弛(线性规划)问题($\tilde{P}$)的基本最优解 $\tilde{x}^*$，并通过 $\tilde{x}^*$，求得一个基本覆盖 $J^*$(见(3.49)与(3.50)). 然后，进行(4).

(4) 若 $cx(J^*)<x_0^*$，则改进记录解，置 $J^*\to J$，$cx(J^*)\to x_0^*$，然后，进行(5)；若 $cx(J^*)\geqslant x_0^*$，则进行(5).

(5) 将矩阵 $\boldsymbol{A}$ 和向量 $\boldsymbol{c}$ 排列成如下形式：

$$\boldsymbol{A}=\begin{bmatrix}\boldsymbol{I}_1 & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix},$$

$$\boldsymbol{c}=(\boldsymbol{c}_{J^*},\boldsymbol{c}_{N^*}).$$

同时，取基阵为

$$\boldsymbol{B}(J^*)=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{O}\\ \boldsymbol{A}_{21} & -\boldsymbol{I}_2\end{pmatrix}.$$

然后，进行(6).

(6) 若对所有的 $j\in N^*$，满足 $\boldsymbol{c}_{J^*}\boldsymbol{P}_j-c_j\leqslant 0$. (其中 $\boldsymbol{P}_j$ 为 $\boldsymbol{A}_{12}$中的列向量)，则步骤终止，当前的记录解 $x(J)$，便是($P$)的最优解. 否则，置

$$Q=\{j\mid \boldsymbol{c}_{J^*}\boldsymbol{P}_j-c_j>0,j\in N^*\}.$$

然后，进行(7).

(7) 增加割平面：

$$\sum_{j\in Q}x_j\geqslant 1.$$

用覆盖问题：

$$\min\{cx\mid Ax\geqslant 1,\sum_{j\in Q}x_j\geqslant 1,x\text{ 为 }0\text{—}1\text{ 向量}\}.$$

代替原来的问题(P)，转到(2).

**例 3.7.3** 求解

$$\min \quad x_0 = 7x_1 + 3x_2 + 7x_3 + 12x_4 + 6x_5,$$

$$\text{s.t.} \quad x_1 + x_3 + x_4 \geqslant 1,$$
$$x_2 + x_4 \geqslant 1,$$
$$x_1 + x_5 \geqslant 1,$$
$$x_3 + x_5 \geqslant 1,$$
$$x_1, \cdots, x_5 \text{ 取 } 0 \text{ 或 } 1.$$

(1) 任取基本覆盖{1,2,3}作为初始的记录 $J$，置 $x_0^* = 7+3+7=17$.

(5) 排列矩阵 $\boldsymbol{A}$ 和向量 $\boldsymbol{c}$ 成如下形式：

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array}$$

$$\boldsymbol{c} = (7 \quad 3 \quad 7 \quad 12 \quad 6)$$

$$\boldsymbol{A} = \left[\begin{array}{ccc:cc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ \hdashline 1 & 0 & 1 & 1 & 0 \end{array}\right] \begin{array}{l} \text{—(原第 3 行)} \\ \text{—(原第 2 行)} \\ \text{—(原第 4 行)} \\ \text{—(原第 1 行)} \end{array},$$

取基矩阵 $\boldsymbol{B}(J)$ 为

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & -1 \end{bmatrix},$$

$$\boldsymbol{A}_{12} = (\boldsymbol{P}_4, \boldsymbol{P}_5) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\boldsymbol{c}_B = (7,3,7,0) = (\boldsymbol{c}_J, 0).$$

(6) $\boldsymbol{c}_J\boldsymbol{P}_4 - c_4 = 3-12 = -9 < 0$,

$\boldsymbol{c}_J\boldsymbol{P}_5 - c_5 = 14-6 = 8 > 0$.

置 $Q = \{5\}$.

(7) 增加条件 $x_5 \geqslant 1$ 后 $\boldsymbol{A}$ 成为

$$\begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array}$$
$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(2) 化简 $\boldsymbol{A}$：

由规则 1，可取 $x_5=1$，划去第 5 列和第 1，3，5 行后得

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array}$$
$$\boldsymbol{A}=\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix},$$

$$\boldsymbol{c}=(7,3,7,12).$$

(3) 松弛(线性规划)问题的最优解为：

$$x_1=1,x_2=1,x_3=0,x_4=0,x_5=1.$$

因此，它也是覆盖问题的最优解. 步骤终止.

## 3.8 目标函数为分式时的整数规划

考虑问题($P$)：

$$\max \quad \lambda=\frac{\boldsymbol{cx}+d}{\boldsymbol{ex}+f}, \tag{3.51}$$

$$\text{s.t.} \quad \boldsymbol{Ax}=\boldsymbol{b}, \boldsymbol{x} \text{ 的分量取非负整数} \tag{3.52}$$

其中，$\boldsymbol{e}$ 为分量取非负的行向量($f>0$).

假如上述规划问题是某一生产计划问题的数学模型，那么，目标函数的分母可表示生产成本，分子可表示总产值，而 $\lambda$ 可表示资金的增长率.

若上述问题的最优解为 $\boldsymbol{x}^*$，目标函数最大值为 $\boldsymbol{\lambda}^*$，则对问题的任何可行解 $\boldsymbol{x}$，有关系：

$$\frac{cx+d}{ex+f} \leqslant \lambda^*,$$

即

$$(cx+d) - \lambda^*(ex+f) \leqslant 0.$$

因此，线性整数规划问题 $P(\lambda^*)$：

$$\max \quad x_0(\lambda^*) = (cx+d) - \lambda^*(ex+f), \tag{3.53}$$

$$\text{s.t.} \quad Ax = b; x \text{ 的分量取非负整数}. \tag{3.54}$$

的最大值为零.

假如，考虑问题 $P(\lambda)$：

$$\max \quad x_0(\lambda) = (cx+d) - \lambda(ex+f), \tag{3.55}$$

$$\text{s.t.} \quad Ax = b, x \text{ 的分量取非负整数}. \tag{3.56}$$

则显然，当 $\lambda<\lambda^*$ 时，$\max x_0(\lambda)>0$；当 $\lambda>\lambda^*$ 时，$\max x_0(\lambda)<0$.

分式整数规划问题($P$)的迭代算法步骤.

它是利用线性整数规划的解来逼近分式整数规划的解.

假设已知问题($P$)的一个可行解 $x^1$，

$$\lambda^1 = \frac{cx^1+d}{ex^1+f}.$$

方法：

(0) 置 $i=1$.

(1) 求解线性整数规划问题 $P(\lambda^i)$：

$$\max \quad x_0(\lambda^i) = (cx+d) - \lambda^i(ex+f),$$

$$\text{s.t.} \quad Ax = b;\ x \text{ 的分量取非负整数}.$$

(2) 若 $\max x_0(\lambda^i)=0$ 则步骤终止. $x^i$ 便是最优解. 相反，设求得 $P(\lambda^i)$的最优解为 $x^{i+1}$.

(3) 计算

$$\lambda^{i+1} = \frac{cx^{i+1}+d}{ex^{i+1}+f}.$$

显然，$\lambda^{i+1}>\lambda^i$.

(4) 置 $i+1 \to i$，然后，转到(1).

因为 $\lambda^1 < \lambda^2 < \cdots\cdots < \lambda^s$，故在迭代过程中，所有的 $x^i$ 各不相同. 因此，只要问题的可行解集合有界，算法必在有限步内终止.

## 3.9 割平面法的新进展

### 3.9.1 常用符号和基本概念

$$R^n = \{x \mid x = (x_1, \cdots, x_n)^{\mathrm{T}}, \forall\ x_j \in Q\}.$$

$$R^n_+ = \{x \mid x \in R^n, x \geqslant 0\}.$$

$$Z^n = \{x \mid x \in R^n, \forall\ x_j \in Z\}.$$

$$Z^n_+ = \{x \mid x \in Z^n, x \geqslant 0\}.$$

$$A = (A_1, \cdots, A_n), A_j \in Z^m, (j = 1, \cdots, n).$$

$$b = (b_1, \cdots, b_m)^{\mathrm{T}},\ b \in Z^m.$$

$$c = (c_1, \cdots, c_n)^{\mathrm{T}},\ c \in Z^n.$$

$$S = \{x \mid x \in Z^n_+, Ax \leqslant b\}.$$

为简明起见，在本节中，把整数线性规划问题：

$$\max \quad c^{\mathrm{T}}x \tag{3.57}$$

$$\text{s.t.} \quad x \in S \tag{3.58}$$

简记如下：

$$\max\{c^{\mathrm{T}}x \mid x \in S\}. \tag{3.59}$$

整数规划中的一个最基本的概念是“$S$-凸包”($S$-convex hull)：

$$\mathrm{conv}(S) = \left\{x = \sum_{i=1}^{t}\lambda_i x^i \mid \sum_{i=1}^{t}\lambda_i = 1, \forall\ \lambda_i \geqslant 0, \forall\ x^i \in S,\ t\ \text{为任意正整数}\right\}. \tag{3.60}$$

即使 $S$ 是无限点集，conv($S$)仍然是一凸多面体，即存在有限个向量：

$$(\pi^i, \pi^i_0)^{\mathrm{T}} \in R^{n+1} \quad (i = 1, \cdots, s)$$

使得：

$$\text{conv}(S)=\{x\in R_+^n \mid \pi^i x\leqslant \pi_0^i, i=1,\cdots,s\}$$

所有的子集：

$$\{x \mid \pi^i x=\pi_0^i, x\in \text{conv}(S)\} \quad (i=1,\cdots,S)$$

构成了凸多面体 conv($S$)的所有边界面(facet). 假如能写出 conv($S$)的不等式条件，那么，只要求得线性规划问题：

$$\max\{c^{\mathrm{T}}x \mid \pi^i x\leqslant \pi_0^i, i=1,\cdots,s, x\geqslant 0\}.$$

的基本最优解，便可得到整数线性规划问题的最优解. 割平面方法的核心问题是研究如何从定义 $S$ 的条件导出 conv($S$)的边界面.

下面，简单介绍一些本节将要用到的图论基本概念.

一个**图**(graph)$G$ 通常用二元组($V$,$E$)来表示，其中 $V$ 表示点的集合，$E$ 表示以一些点对为元素所构成的集合，称为边的集合. 对任一边 $e=[v_i,v_j]\in E$，称 $e$ 关联于点 $v_i$ 和 $v_j$；也称点 $v_i$ 和 $v_j$ 相邻. 若两条不同的边关联于同一个点，则称这两条边相邻. $V$ 的一个子集 $K$，若其中的点都两两相邻，则称 $K$ 为图 $G$ 的一个**团**(clique)；设 $K$ 是图 $G$ 的一个团，若 $G$ 中不存在包含 $K$ 的更大的团，则称 $K$ 是 $G$ 的一个**极大团**(maximal clique). $V$(或 $E$)的一个子集 $S$，若其中的元素两两互不相邻，则称 $S$ 为图的一点(或边)**无关集**. 设 $V=\{v_1,\cdots,v_n\}$，$E=\{e_1,\cdots,e_m\}$，对 $V$(或 $E$)的任一子集 $S$，定义 0—1 向量 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$(或$(x_1,\cdots,x_m)^{\mathrm{T}}$)，使得：

$$x_j=\begin{cases}1, & \text{当 } v_j\in S(\text{或 } e_j\in S) \text{ 时},\\ 0, & \text{当 } v_j\notin S(\text{或 } e_j\notin S) \text{ 时},\end{cases}$$

则称 $x$ 是点集 $S$(或边集 $S$)的**关联向量**(incidence vector). 一个 $n\times m$ 的 0—1 矩阵 $A=(a_{ij})$，若使：

$$a_{ij}=\begin{cases}1, & \text{当 } v_i \text{ 关联于边 } e_j \text{ 时},\\ 0, & \text{相反},\end{cases}$$

则称 $A$ 是图 $G$ 的点-边**关联矩阵**(incidence matrix of a graph). 一

个点的序列：$P=\{v_{i_1},v_{i_2},\cdots,v_{i_k}\}$，若使：$(v_{i_1},v_{i_2})$，$(v_{i_2},v_{i_3})$，…，$(v_{i_{k-1}},v_{i_k})$都是图的边，则称 $P$ 是图中的一条连结 $v_{i_1}$ 和 $v_{i_k}$ 的**链**(chain)．凡序列中的点不重复出现的链，称为**简单链**(simple chain)．若图 $G$ 中任何两点间都有链相连结，则称 $G$ 是一个**连通图**(connected graph)．一条链 $P=\{v_{i_1},v_{i_2},\cdots,v_{i_k}\}$，若使 $v_{i_1}=v_{i_k}$，则称 $P$ 是图中的一个**圈**(circuit)；若同时又使$\{v_{i_1},\cdots,v_{i_{k-1}}\}$成为一条简单链，则称 $P$ 是一个**简单圈**(simple circuit)．包含图中所有点的简单圈，称为**哈密尔顿圈**(Hamilton circuit)．一个网络 $N$ 通常用三元组$(V,E,w)$来表示，其中$(V,E)$确定了一个图，$w$ 是定义在 $E$ 上的一个非负函数．对任意的 $e\in E$，称函数值 $w(e)$为边 $e$ 的**容量**(capacity)．对任一点的真子集 $S$，令 $\overline{S}=V\backslash S$．定义：

$$\delta(S)=\{(v_i,v_j)\mid(v_i,v_j)\in E,v_i\in S,v_j\in\overline{S}\}. \quad (3.61)$$

$$w(\delta(S))=\sum_{e\in\delta(S)}w(e). \quad (3.62)$$

称 $\delta(S)$为网络的一个**割集**(cut-set)，称 $w(\delta(S))$为割集 $\delta(S)$的容量．使容量达到最小的割集，称为网络的**最小割集**(minimal cut-set)．对任意的矩阵 $\boldsymbol{A}=(\boldsymbol{A}_1,\cdots,\boldsymbol{A}_n)$，定义图 $G[\boldsymbol{A}]=(V,E)$如下：

$$V=\{1,\cdots,n\},$$

$$E=\{(i,j)\mid\boldsymbol{A}_i^{\mathrm{T}}\boldsymbol{A}_j>0,i,j\in V\}.$$

称 $G[\boldsymbol{A}]$为矩阵 $\boldsymbol{A}$ 的**交图**(intersection graph)．设 $G=(V,E)$，$E=\{e_1,\cdots,e_m\}$．若向量 $\boldsymbol{x}=(x_1,\cdots,x_m)^{\mathrm{T}}$，则有时记 $x_j=x(e_j)$，即$x(e_j)$为对应于 $e_j$ 的 $\boldsymbol{x}$ 的分量．

### 3.9.2 分离不等式(分离面)

一个不等式：$\boldsymbol{\pi x}\leqslant\pi_0$，若满足：$\boldsymbol{x}\in\mathrm{conv}(S)\Rightarrow\boldsymbol{\pi x}\leqslant\pi_0$，则称其为 $S$ 的一个**分离不等式**(valid inequality)或称**分离面**(face)．下面将介绍几种导出分离面的基本方法．

### 3.9.3 Gomory-Chvatal 法

对任意的 $\boldsymbol{u}=(u_1,\cdots,u_m)\geqslant \boldsymbol{0}$，有：

$$\sum_{j=1}^{n} \boldsymbol{u}\boldsymbol{A}_j \boldsymbol{x} \leqslant \boldsymbol{u}\boldsymbol{b}.$$

因为条件 $\boldsymbol{x}\geqslant \boldsymbol{0}$，含有：

$$\sum_{j=1}^{n} (\boldsymbol{u}\boldsymbol{A}_j - \lfloor \boldsymbol{u}\boldsymbol{A}_j \rfloor)\boldsymbol{x}_j \geqslant 0.$$

因此有关系式

$$\sum_{j=1}^{n} (\lfloor \boldsymbol{u}\boldsymbol{A}_j \rfloor) x_j \leqslant \boldsymbol{u}\boldsymbol{b}.$$

因为条件 $\boldsymbol{x}\in Z^n$. 含有：

$$\sum_{j=1}^{n} (\lfloor \boldsymbol{u}\boldsymbol{A}_j \rfloor) x_j \leqslant \lfloor \boldsymbol{u}\boldsymbol{b} \rfloor. \tag{3.63}$$

假如所选的 $\boldsymbol{u}\geqslant \boldsymbol{0}$ 满足：

$$\boldsymbol{u}\boldsymbol{b} - \lfloor \boldsymbol{u}\boldsymbol{b} \rfloor > 0, \boldsymbol{u}\boldsymbol{A}_j = \lfloor \boldsymbol{u}\boldsymbol{A}_j \rfloor (j = 1,\cdots,n).$$

则就是通常所称的割平面.

### 3.9.4 同余法(modulo method)

设

$$S = \{\boldsymbol{x} \mid \boldsymbol{x} \in Z_+^n, \sum_{j=1}^{n} a_j x_j = a_0\}, \forall\ a_j \in R^1.$$

设 $d$ 是任意给定的正整数. 考虑：

$$S_d = \{\boldsymbol{x} \mid \boldsymbol{x} \in Z_+^n, k \in Z^1, \sum_{j=1}^{n} a_j x_j - kd = a_0\},$$

显然 $S\subseteq S_d$. 设

$$a_j = b_j + \alpha_j d, 0 \leqslant b_j < d, \alpha_j \in Z^1, j = 0,1,\cdots,n.$$

则 $$S_d = \{\boldsymbol{x} \mid \boldsymbol{x} \in Z_+^n, \sum_{j=1}^{n} b_j x_j = b_0 + yd, y \in Z^1\}.$$

因为

$$\sum_{j=1}^{n} b_j x_j \geqslant 0, 0 \leqslant b_0 < d, y \in Z^1,$$

故对任意的 $\boldsymbol{x} \in S_d$，必使 $y \geqslant 0$，因此

$$\boldsymbol{x} \in S \subseteq S_d \Rightarrow \sum_{j=1}^{n} b_j x_j \geqslant b_0.$$

特别的，当 $a_0$ 不是整数时，可取 $d=1$，这时

$$b_j = a_j - \lfloor a_j \rfloor \quad (j = 0, 1, \cdots, n).$$

因此，上述条件可写为：

$$\sum_{j=1}^{n} (a_j - \lfloor a_j \rfloor) x_j \geqslant a_0 - \lfloor a_0 \rfloor.$$

这就是大家所熟悉的 Gomory 分数割平面.

### 3.9.5 逻辑和法(logical summation method)

设 $\boldsymbol{\pi}^1 \boldsymbol{x} \leqslant \pi_0^1$ 是 $S_1(\subset R_+^n)$的一个分离面；$\boldsymbol{\pi}^2 \boldsymbol{x} \leqslant \pi_0^2$ 是 $S_2(\subset R_+^n)$的一个分离面，则显然

$$\sum_{j=1}^{n} \min(\pi_j^1, \pi_j^2) x_j \leqslant \max(\pi_0^1, \pi_0^2)$$

(其中 $\boldsymbol{\pi}^i = (\pi_1^i, \cdots, \pi_n^i)$)是 $S_1 \cup S_2$ 的一个分离面.

设

$$S = \left\{ \begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix} \middle| x_0 \in Z^1, \boldsymbol{x} \in R_+^n, x_0 = a_0 - \sum_{j=1}^{n} a_j x_j \right\},$$

$$S_1 = \left\{ \begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix} \middle| \begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix} \in S, x_0 \leqslant \lfloor a_0 \rfloor \right\},$$

$$S_2 = \left\{ \begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix} \middle| \begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix} \in S, x_0 \geqslant \lfloor a_0 \rfloor + 1 \right\},$$

其中$\begin{bmatrix} x_0 \\ \boldsymbol{x} \end{bmatrix}$表示 $n+1$ 维列向量.

则

$$S = S_1 \cup S_2.$$

$$S_1 = \left\{x \in R_+^n \;\middle|\; \sum_{j=1}^{n} a_j x_j \geqslant a_0 - \lfloor a_0 \rfloor\right\}$$

$$= \left\{x \in R_+^n \;\middle|\; \frac{1}{\lfloor a_0 \rfloor - a_0} \sum_{j=1}^{n} a_j x_j \leqslant -1\right\},$$

$$S_2 = \left\{x \in R_+^n \;\middle|\; \sum_{j=1}^{n} a_j x_j \leqslant a_0 - \lfloor a_0 \rfloor - 1\right\}$$

$$= \left\{x \in R_+^n \;\middle|\; \frac{1}{\lfloor a_0 \rfloor + 1 - a_0} \sum_{j=1}^{n} a_j x_j \leqslant -1\right\}.$$

让

$$\pi_j = \min\left(\frac{1}{\lfloor a_0 \rfloor - a_0} a_j, \frac{1}{\lfloor a_0 \rfloor + 1 - a_0} a_j\right)$$
$$(j = 1, \cdots, n).$$

则

$$\sum_{j=1}^{n} \pi_j x_j \leqslant -1$$

是 $S$ 的一个分离面.

设

$$\bar{P} = \{x \in R_+^n \mid Ax \leqslant b, \forall\ x_j \leqslant 1\},$$
$$B^n = \{x \in Z_+^n \mid x_j \leqslant 1, j = 1, \cdots, n\},$$
$$\bar{S} = \bar{P} \cap B^n.$$

则有

**性质 3.9.1** conv($\bar{S}$)的边界面是形式为(3.63)的 Gomory-Chvatal 分离面.

### 3.9.6 (SA)函数法(superadditive function)

函数 $f: R^m \to R^1$,若满足条件:

$$\begin{cases} f(\mathbf{0}) = 0, \\ f(\boldsymbol{d}_1) + f(\boldsymbol{d}_2) \leqslant f(\boldsymbol{d}_1 + \boldsymbol{d}_2), \forall\ \boldsymbol{d}_1, \boldsymbol{d}_2 \in R^m, \end{cases}$$

则称 $f$ 是(SA)函数.

函数 $f$: $R^m \to R^1$,若满足条件:

$$f(\boldsymbol{d}_1) \leqslant f(\boldsymbol{d}_2), \forall\ \boldsymbol{d}_1 \leqslant \boldsymbol{d}_2 \in R^m,$$

则称 $f$ 是**非降函数**(nondecreasing function). 容易验证,对任意给定的 $\boldsymbol{u}^{\mathrm{T}} \in R_+^m$,函数 $f(\boldsymbol{d}) = [\boldsymbol{u}\boldsymbol{d}]$是一非降的(SA)函数.

**性质 3.9.2** 设 $f$: $R^m \to R^1$ 是一个非降的(SA)函数,$\boldsymbol{A}$ 是 $m \times n$ 的有理数矩阵,$\boldsymbol{b} \in R^m$,

$$\boldsymbol{A} = (\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n), \text{设}$$
$$S = Z_+^n \cap \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \leqslant \boldsymbol{b}\},$$

则

$$\sum_{j=1}^{n} f(\boldsymbol{A}_j) x_j \leqslant f(\boldsymbol{b}) \tag{3.64}$$

是 $S$ 的一个分离面.

非降(SA)函数的例子:

设

$$\begin{cases} f_\alpha: R^1 \to R^1 \quad (0 \leqslant \alpha < 1), \\ f_\alpha(d) = \begin{cases} \lfloor d \rfloor, & \text{若 } d - \lfloor d \rfloor \leqslant \alpha, \\ \lfloor d \rfloor + \dfrac{d - \lfloor d \rfloor - \alpha}{1 - \alpha}, & \text{若 } d - \lfloor d \rfloor > \alpha. \end{cases} \end{cases}$$

则对任意的 $\alpha(0 \leqslant \alpha < 1)$,函数:

$$f_\alpha(d) = \lfloor d \rfloor + \frac{1}{1-\alpha}(d - \lfloor d \rfloor - \alpha)^+$$

都是非降的(SA)函数. 而函数:

$$f_\alpha(\lfloor \boldsymbol{u}\boldsymbol{d} \rfloor): R^m \to R^1 (u^T \in R_+^m, d \in R^m),$$

对任意给定的 $\alpha$ 和 $\boldsymbol{u}$,也是一个非降的(SA)函数.

**性质 3.9.3** conv($S$)的边界面是形式为(3.64)的(SA)函数分离面.

### 3.9.7 升维法(lifting method)

下面介绍一种从多面体的低维边界面构成较高维边界面的升维方法.

设

$$P=\{x\in R_+^n \mid Ax\leqslant b\};$$
$$\overline{P}=\{x\mid \forall\ x_j\leqslant 1\}\cap P;$$
$$\overline{S}=\overline{P}\cap Z^n;$$
$$\overline{S}^0=\overline{S}\cap\{x\in B^n\mid x_1=0\};$$
$$\overline{S}^1=\overline{S}\cap\{x\in B^n\mid x_1=1\}.$$

**性质 3.9.4** 若 $\sum_{j=2}^{n}\pi_j x_j\leqslant\pi_0$ 是 $\mathrm{conv}(\overline{S}^0)$ 的一个边界面,且

$$\zeta=\max\left\{\sum_{j=2}^{n}\pi_j x_j \mid x\in\overline{S}^1\right\},$$

则

$$(\pi_0-\zeta)x_1+\sum_{j=2}^{n}\pi_j x_j\leqslant\pi_0$$

是 $\mathrm{conv}(\overline{S})$ 的一个边界面.

**性质 3.9.5** 若 $\sum_{j=2}^{n}\pi_j x_j\leqslant\pi_0$ 是 $\mathrm{conv}(\overline{S}^1)$ 的一个边界面,且

$$\zeta=\max\left\{\sum_{j=2}^{n}\pi_j x_j \mid x\in\overline{S}^0\right\},$$

则

$$(\zeta-\pi_0)x_1+\sum_{j=2}^{n}\pi_j x_j\leqslant\pi_0+(\zeta-\pi_0)$$

是 $\mathrm{conv}(\overline{S})$ 的一个边界面.

### 3.9.8 装箱(packing)多面体的边界面

设 $A=(A_1,\cdots,A_n)$ 是 $m\times n$ 的 0—1 矩阵,$\mathbf{1}=(1,\cdots,1)^{\mathrm{T}}\in$

$R^m$. 考虑集合：

$$P_A = \{x \in B^n \mid Ax \leqslant 1\}.$$

通常称其为装箱多面体. 定义 $A$ 的交图 $G[A]=(V,E)$ 如下：

$$\begin{cases} V = \{1,2,\cdots,n\}, \\ E = \{[i,j] \mid A_i^T A_j \geqslant 1, i,j \in V\}, \end{cases}$$

则容易看出 $x \in P_A \Leftrightarrow x$ 是 $G[A]$中某点无关集的关联向量.

因此，对 $G[A]$中的任何一个团 $K$，必有：

$$x \in P_A \Rightarrow \sum_{j \in K} x_j \leqslant 1.$$

**性质 3.9.6** 若 $K$ 是 $G[A]$的极大团，则条件 $\sum_{j \in K} x_j \leqslant 1$ 是 $\mathrm{conv}(P_A)$的边界面.

### 3.9.9 背包问题的边界面

考虑(0,1)背包问题

$$S = \{x \in B^n \mid \sum_{j \in N} a_j x_j \leqslant a_0\},$$

其中 $N=\{1,\cdots,n\}$.

不妨可设 $a_0 \geqslant a_1 \geqslant a_2 \geqslant \cdots \geqslant a_n > 0$ 对任意的子集 $R \subseteq N$，让 $x^R$ 表示 $R$ 的特征向量(或关联向量)，即 $x^R=(x_1^R,\cdots,x_n^R)^T$，

$$x_j^R = \begin{cases} 1, & j \in R, \\ 0, & j \notin R. \end{cases}$$

若 $x^R \in S$，则称 $R$ 是一个**独立集**；若 $x^R \notin S$，则称 $R$ 是一个**相关集**(dependent set). 若 $C$ 是一个相关集，而对任意的 $j \in C$，$C \setminus \{j\}$都是独立集，则称 $C$ 是一个**极小相关集**(minimal dependent set).

**性质 3.9.7** 设 $C=\{j_1,\cdots,j_r\}$是一个极小相关集，$j_1<j_2<\cdots<j_r$；设 $E(C)=C\cup\{1,2,\cdots,j_1\}$；若 $N\setminus E(C) \neq \varnothing$，且$(C\setminus\{j_1,j_2\})\cup\{1\}$和$(C\setminus\{j_1\})\cup\{p\}$都是独立集，其中 $p=\min\{j \mid j \in N\setminus E(C)\}$；或者，若 $N\setminus E(C)=\varnothing$，而$(C\setminus\{j_1,j_2\})\cup\{1\}$是独立集，则

条件：

$$\sum_{j\in E(C)} x_j \leqslant |C| - 1$$

是 conv($S$)的一个边界面.

### 3.9.10 匹配多面体(matching polyhedron)

对图 $G=(V,E)$，一个边子集 $X\subseteq E$，若使图中每个点只关联于 $X$ 中的一条边，则称 $X$ 是 $G$ 的一个匹配. 一个边子集 $X\subseteq E$，若使图中每个点只关联于 $X$ 中的两条边，则称 $X$ 是 $G$ 的一个 2-匹配(2-Matching). 我们用 $\boldsymbol{x}$ 表示边子集也表示它的关联向量，用 $X(e)$表示 $\boldsymbol{x}$ 的分量，其中 $e$ 是一条边. 让 $A$ 表示 $G$ 的点边关联矩阵. 则

匹配问题可表示为如下形式：

$$\{\boldsymbol{x}\,|\,A\boldsymbol{x} = \boldsymbol{1}, \boldsymbol{x} \geqslant \boldsymbol{0}, \boldsymbol{x}\text{ 是整数向量}\}.$$

2-匹配问题可表示为如下形式：

$$\{\boldsymbol{x}\,|\,A\boldsymbol{x} = 2\boldsymbol{1}, \boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x}\text{ 是 } 0,1 \text{ 向量}\}.$$

一个点子集 $T\subseteq V$，若它所含的点数$|T|$是奇数，则称 $T$ 为奇点集. 定义

$$E(T) = \{e \in E\,|\,e = [v_j, v_k], v_j \in T, v_k \in T\},$$

$$\delta(T) = \{e \in E\,|\,e = [v_i, v_k], v_i \in T, v_k \notin T\}.$$

**性质 3.9.8** 图 $G$ 的匹配的凸包为：

$$\sum_{e\in\delta(v)} x(e) = 1, \quad \forall v \in V, \tag{3.65}$$

$$\sum_{e\in\delta(T)} x(e) \geqslant 1, \quad \forall \text{奇点集 } T \subset V, \tag{3.66}$$

$$x(e) \geqslant 0, \quad \forall e \in E. \tag{3.67}$$

这就是著名的 Edmonds-匹配多面体(Edmonds matching polyhedron).

设 $\boldsymbol{x}$ 是任意 2-匹配的关联向量，则对任意的子集 $S\subset V$，以及

$F\subseteq\delta(S)$，有关系

$$\begin{cases} \sum_{e\in\delta(S)} x(e) \geqslant \sum_{e\in F} x(e), \\ |F| \geqslant \sum_{e\in F} x(e). \end{cases}$$

因此

$$\sum_{e\in\delta(S)} x(e) + |F| \geqslant 2\sum_{e\in F} x(e).$$

另一方面，因为

$$2\sum_{e\in E(S)} x(e) + \sum_{e\in\delta(S)} x(e) = 2|S|.$$

因此，数值 $\sum_{e\in\delta(S)} x(e)$ 是偶数. 现在，若$|F|$是奇数，则有

$$\sum_{e\in\delta(S)} x(e) + |F| \geqslant 2\sum_{e\in F} x(e) + 1,$$

即

$$2|S| - 2\sum_{e\in E(S)} x(e) + |F| \geqslant 2\sum_{e\in F} x(e) + 1.$$

因而有

$$\sum_{e\in E(S)} x(e) + \sum_{e\in F} x(e) \leqslant |S| + \left[\frac{|F|-1}{2}\right]. \qquad (3.68)$$

通常称二元组"$S,F$"($S\subset V,F\subseteq\delta(S)$，$|F|$是奇数)为一个"梳子"

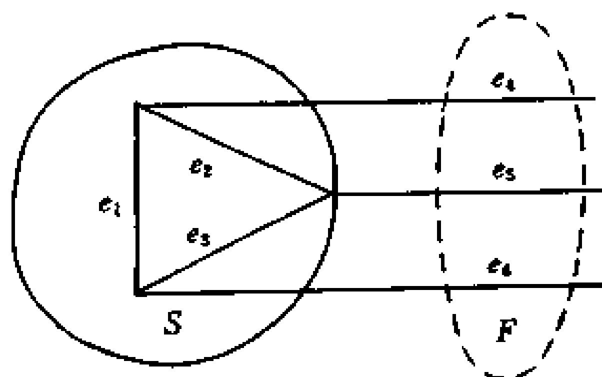



图 3.9

(comb),称关系式(3.68)为一个梳子不等式.最简单的梳子不等式为如下形式

$$\sum_{j=1}^{6} x(e_j) \leqslant 4.$$

称其为**"三角形梳子"**(triangular comb)(因 $S$ 中点形成的子图为三角形).

**性质 3.9.9** 图 $G$ 的 2-匹配的凸包为

$$\begin{cases} \sum_{e\in\delta(v)} x(e) = 2, \forall\, v \in V; \\ \sum_{e\in E(S)} x(e) + \sum_{e\in F} x(e) \leqslant |S| + \left[\frac{|F|-1}{2}\right], \forall \text{ 梳子“}S,F\text{”}; \\ \quad x(e) \geqslant 0, \quad \forall\, e \in E; \\ \quad x(e) \leqslant 1, \quad \forall e \in E. \end{cases}$$

### 3.9.11 Hamilton 圈(Hamilton circuit)

包含所有点的简单圈(点不重复的圈)称为 Hamilton 圈.或者说,关联于每个点的边数为 2 的连通图.用向量 $x$ 表示 Hamilton 圈的关联向量,则它的数学形式为:

$$\begin{cases} \sum_{e\in\delta(v)} x(e) = 2, \quad \forall\, \nu \in V; \\ \sum_{e\in\delta(B)} x(e) \geqslant 2, \quad \forall\, B \subset V, \varnothing \neq B \neq V; \\ x(e) \leqslant 1, \quad \forall\, e \in E; \\ x(e) \geqslant 0, \quad \forall\, e \in E; \\ x(e) \text{ 取整数值}, \forall\, e \in E. \end{cases}$$

第一组条件表示关联于每个点的边数为 2;第二组条件表示每个割集中至少含有两条边.利用 2-匹配凸包,可得 Hamilton 圈的松弛问题如下:

$$\begin{cases} \sum_{e \in \delta(v)} x(e) = 2, & \forall v \in V; \\ \sum_{e \in E(S)} x(e) + \sum_{e \in F} x(e) \leqslant |S| + \left[\frac{|F|-1}{2}\right], & \forall 梳子“S,F”; \\ \sum_{e \in \delta(B)} x(e) \geqslant 2, & \forall B \subset V, \varnothing \neq B \neq V; \\ x(e) \leqslant 1, & \forall e \in E; \\ x(e) \geqslant 0, & \forall e \in E. \end{cases}$$

求 Hamilton 圈的割平面算法步骤：

设 $Q$ 是任给的某些梳子的集合，$R$ 是由图的某些点子集所构成的集合．设 $\boldsymbol{x}$ 是下述问题的一个解：

$$\begin{cases} \sum_{e \in \delta(v)} x(e) = 2, \quad \forall v \in V; & (3.69) \\ \sum_{e \in E(S)} x(e) + \sum_{e \in F} x(e) \leqslant |S| + \left[\frac{|F|-1}{2}\right], & \\ \qquad\qquad “S,F” \in Q; & (3.70) \\ \sum_{e \in \delta(B)} x(e) \geqslant 2, \quad B \in R; & (3.71) \\ x(e) \leqslant 1, \quad \forall e \in E; & (3.72) \\ x(e) \geqslant 0, \quad \forall e \in E. & (3.73) \end{cases}$$

若 $x$ 是整数解，则步骤终止，否则进行步骤(1)．

(1) 构造以 $x(e)$ 为边 $e$ 的容量的网络 $G(x)$．

(2) 用 Gomory—T. C. Hu 方法，求网络 $G(x)$ 的**最小割集**(min-cut set)$\delta(Y)$．

(3) 若

$$\sum_{e \in \delta(Y)} x(e) < 2, 则 Y \cup R \to R.$$

若

$$\max_{B \in R} \sum_{e \in \delta(B)} x(e) = \sum_{e \in \delta(B^*)} x(e) > 2,$$

则 $$R\backslash\{B^*\}\to R.$$
重新寻求问题(3.69)～(3.73)的解 $x$,然后转到(1).

(4) 若

$$\sum_{e\in\delta(Y)} x(e)\geqslant 2,$$

则检查各个梳子不等式(首先检查最简单的三角梳子不等式).

若有某个梳子"$S,F$",使(3.70)关系式不满足,则"$S,F$"$\cup Q\to Q$,重新寻求问题(3.69)～(3.73)的解 $x$.

若 $x$ 不是整数解,则转到(1).

检验一般的梳子不等式是很复杂的.它要利用求最小奇点割集的算法,要将 2-匹配问题转化到匹配问题.但是,确已找到了一种检验所有梳子不等式的多项式算法.

# 参考文献


1 Balas E. An additive algorithm for solving linear programs with zero-one variables. Operations Reseach 13 1965 517～546

2 Dantzig G, Fulkerson R and Johnson S, Solution of a large-scale traveling-salesman problem. Operations Research 2 1954 393～410

3 Edmonds J and Giles R. A min-max relation for submodular functions on graphs. in: Studies in Integer Programming (P. L. Hammer, et al. eds.) Annals of Discrete Mathematics 1 1977 185～204

4 Fulkerson D R. Blocking and anti-blocking pairs of polyhedra, Mathematical Programming 1 1971 168～194

5 Garfinkel R S and Nemhauser G L. Integer Programming. Wiley, New York, 1972

6 Geoffrion A M and Marsten R E. Integer programming algorithms: a framework and state-of-the-art survey. Management Science 18 1971-2 465～491

7 Gomory R E. Outline of an algorithm for integer solutions to linear pro-

grams. Bulletin of the American Mathematical Society 64 1958 275～278

8 Gomory R E. An algorithm for integer solutions to linear programs. in: Recent Advances in Mathematical Programming (R. L. Graves and P. Wolfe, eds.) McGraw Hill, New York, 1963 269～302

9 Grötschel M and Padberg M W. On the symmetric travelling salesman. Mathematical Programming 1979 16 265～302

10 Hu T C. Integer Programming and Network Flows. Addison-Wesley Reading, Mass, 1969

11 Land A H and Doig A G. An automatic method of solving discrete programming problems. Econometrica 1960 28 497～520

12 Lenstra Jr H W. Integer Programming with a fixed number of variables. Mathematics of Operations Research 1983 8 538～548

13 Nemhauser G L and Wolsey L A. Integer and Combinatorial Optimization. Wiley, New York 1987

14 Schrijver A. Theory of linear and integer programming. John Wiley & Sons, 1986

# 4 动态规划

## 4.1 引言

### 4.1.1 动态规划的发展及研究内容

**动态规划**(dynamic programming)是运筹学的一个分支，是求解**决策过程**(decision process)最优化的数学方法. 20 世纪 50 年代初美国数学家 R. E. Bellman 等人在研究**多阶段决策过程**(multistep decision process)的优化问题时，提出了著名的**最优性原理**(principle of optimality)，把多阶段过程转化为一系列单阶段问题，逐个求解，创立了解决这类过程优化问题的新方法——动态规划. 1957 年出版了他的名著《Dynamic Programming》，这是该领域的第一本著作.

动态规划问世以来，在经济管理、生产调度、工程技术和最优控制等方面得到了广泛的应用. 例如最短路线、库存管理、资源分配、设备更新、排序、装载等问题，用动态规划方法比用其它方法求解更为方便.

虽然动态规划主要用于求解以时间划分阶段的动态过程的优化问题，但是一些与时间无关的静态规划(如线性规划、非线性规划)，只要人为地引进时间因素，把它视为多阶段决策过程，也可以用动态规划方法方便地求解.

下面是多阶段决策过程最优化问题的典型例子.

**例 4.1.1** 最短路线问题

下面是一个线路网，连线上的数字表示两点之间的距离(或费

用). 试寻找一条由 $A$ 到 $G$ 距离最短(或费用最省)的路线.

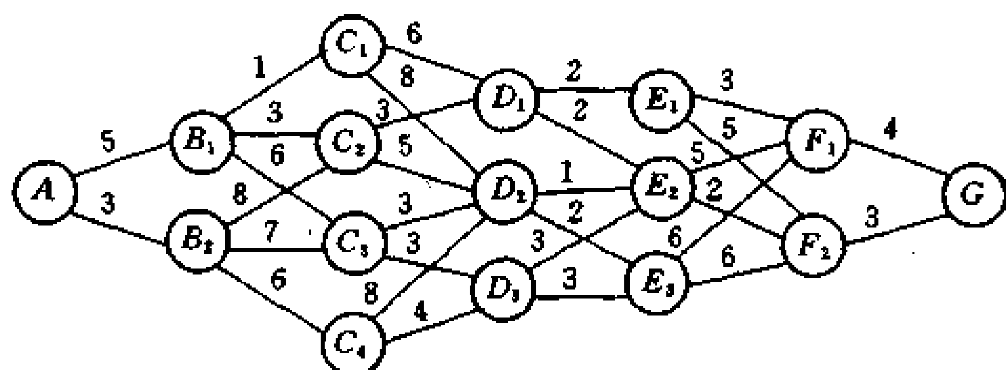



图 4.1

**例 4.1.2** 生产计划问题

工厂生产某种产品,每单位(千件)的成本为 1(千元),每次开工的固定成本为 3(千元),工厂每季度的最大生产能力为 6(千件). 经调查,市场对该产品的需求量第一、二、三、四季度分别为 2,3,2,4(千件). 如果工厂在第一、二季度将全年的需求都生产出来,自然可以降低成本(少付固定成本费),但是对于第三、四季度才能上市的产品需付存储费,每季每千件的存储费为 0.5(千元). 还规定年初和年末这种产品均无库存. 试制订一个生产计划,即安排每个季度的产量,使一年的总费用(生产成本和存储费)最少.

### 4.1.2 决策过程的分类

根据过程的时间变量是离散的还是连续的,分为**离散时间决策过程**(discrete-time decision process),即多阶段决策过程和**连续时间决策过程**(continuous-time decision process);根据过程的演变是确定的还是随机的,分为**确定性决策过程**(deterministic decision process)和**随机性决策过程**(stochastic decision process),其中应用最广的是确定性多阶段决策过程. 4.2~4.4 就是这类过程.

## 4.2 基本概念、基本方程和计算方法

### 4.2.1 动态规划的基本概念

一个多阶段决策过程最优化问题的动态规划模型通常包含以下要素.

**1 阶段**

**阶段**(step)是对整个过程的自然划分.通常根据时间顺序或空间特征来划分阶段,以便按阶段的次序解优化问题.阶段变量一般用 $k=1,2,\cdots,n$ 表示.在例 4.1.1 中由 $A$ 出发为 $k=1$,由 $B_i(i=1,2)$出发为 $k=2$,依此下去从 $F_i(i=1,2)$出发为 $k=6$,共 $n=6$ 个阶段.在例 4.1.2 中按照第一、二、三、四季度分为 $k=1,2,3,4$,共 4 个阶段.

**2 状态**

**状态**(state)表示每个阶段开始时过程所处的自然状况.它应能描述过程的特征并且具有**无后效性**,即当某阶段的状态给定时,这个阶段以后过程的演变与该阶段以前各阶段的状态无关.通常还要求状态是直接或间接可以观测的.

描述状态的变量称**状态变量**(state variable).变量允许取值的范围称**允许状态集合**(set of admissible states).用 $x_k$ 表示第 $k$ 阶段的状态变量,它可以是一个数或一个向量.用 $X_k$ 表示第 $k$ 阶段的允许状态集合.在例 4.1.1 中 $x_2$ 可取 $B_1,B_2$,或将 $B_i$ 定义为 $i$ $(i=1,2)$,则 $x_2=1$ 或 2,而 $X_2=\{1,2\}$.

$n$ 个阶段的决策过程有 $n+1$ 个状态变量,$x_{n+1}$表示 $x_n$ 演变的结果,在例 4.1.1 中 $x_7$ 取 $G$,或定义为 1,即 $x_7=1$.

根据过程演变的具体情况,状态变量可以是离散的或连续的.为了计算的方便有时将连续变量离散化;为了分析的方便有时又

将离散变量视为连续的.

状态变量简称为状态.

**3　决策**

当一个阶段的状态确定后,可以作出各种选择从而演变到下一阶段的某个状态,这种选择手段称为**决策**(decision),在最优控制问题中也称为**控制**(control).

描述决策的变量称**决策变量**(decision variable).变量允许取值的范围称**允许决策集合**(set of admissible decisions).用 $u_k(x_k)$ 表示第 $k$ 阶段处于状态 $x_k$ 时的决策变量,它是 $x_k$ 的函数,用 $U_k(x_k)$表示 $x_k$ 的允许决策集合.在例 4.1.1 中 $u_2(B_1)$可取 $C_1$,$C_2$ 或 $C_3$,可记作 $u_2(1)=1,2,3$,而 $U_2(1)=\{1,2,3\}$.

决策变量简称决策.

**4　策略**

决策组成的序列称为**策略**(policy).由初始状态 $x_1$ 开始的全过程的策略记作 $p_{1n}(x_1)$,即 $p_{1n}(x_1)=\{u_1(x_1),u_2(x_2),\cdots,u_n(x_n)\}$.由第 $k$ 阶段的状态 $x_k$ 开始到终止状态的后部子过程的策略记作 $p_{kn}(x_k)$,即 $p_{kn}(x_k)=\{u_k(x_k),\cdots,u_n(x_n)\}$,$k=2,\cdots,n-1$.类似地,由第 $k$ 到第 $j$ 阶段的子过程的策略记作 $p_{kj}(x_k)=\{u_k(x_k),\cdots,u_j(x_j)\}$.可供选择的策略有一定的范围,称为**允许策略集合**(set of admissible policies),用 $P_{1n}(x_1)$,$P_{kn}(x_k)$,$P_{kj}(x_k)$表示.

**5　状态转移方程**

在确定性过程中,一旦某阶段的状态和决策为已知,下阶段的状态便完全确定.用**状态转移方程**(equation of state transition)表示这种演变规律,写作

$$x_{k+1}=T_k(x_k,u_k),k=1,2,\cdots,n. \tag{4.1}$$

在例 4.1.1 中状态转移方程为 $x_{k+1}=u_k(x_k)$.

**6　指标函数和最优值函数**

**指标函数**(objective function)是衡量过程优劣的数量指标，它是定义在全过程和所有后部子过程上的数量函数，用 $V_{kn}(x_k, u_k, x_{k+1}, \cdots, x_{n+1})$表示，$k=1,2,\cdots,n$. 指标函数应具有可分离性，即 $V_{kn}$可表为 $x_k, u_k, V_{k+1\,n}$的函数，记为

$$V_{kn}(x_k, u_k, x_{k+1}, \cdots, x_{n+1}) = \varphi_k(x_k, u_k, V_{k+1\,n}(x_{k+1}, \cdots, x_{n+1})) \tag{4.2}$$

并且函数 $\varphi_k$ 对于变量 $V_{k+1\,n}$是严格单调的.

过程在第 $j$ 阶段的阶段指标取决于状态 $x_j$ 和决策 $u_j$，用 $v_j(x_j, u_j)$表示. 指标函数由 $v_j(j=1,2,\cdots,n)$组成，常见的形式有

阶段指标之和，即

$$V_{kn}(x_k, u_k, \cdots, x_{n+1}) = \sum_{j=k}^{n} v_j(x_j, u_j), \tag{4.3}$$

阶段指标之积，即

$$V_{kn}(x_k, u_k, \cdots, x_{n+1}) = \prod_{j=k}^{n} v_j(x_j, u_j), \tag{4.4}$$

阶段指标之极大(或极小)，即

$$V_{kn}(x_k, u_k, \cdots, x_{n+1}) = \max_{k\leqslant j\leqslant n}(\min) v_j(x_j, u_j). \tag{4.5}$$

这些形式下第 $k$ 到第 $j$ 阶段子过程的指标函数为 $V_{kj}(x_k, u_k, \cdots, x_{j+1})$.

根据状态转移方程指标函数 $V_{kn}$还可以表示为状态 $x_k$ 和策略 $p_{kn}$的函数，即 $V_{kn}(x_k, p_{kn})$. 在 $x_k$ 给定时指标函数 $V_{kn}$对 $p_{kn}$的最优值称为**最优值函数**(optimal value function)，记作 $f_k(x_k)$，即

$$f_k(x_k) = \underset{p_{kn}\in P_{kn}(x_k)}{\text{opt}} V_{kn}(x_k, p_{kn}), \tag{4.6}$$

其中 opt 可根据具体情况取 max 或 min.

**7 最优策略和最优轨线**

使指标函数 $V_{kn}$达到最优值的策略是从 $k$ 开始的后部子过程的最优策略，记作 $p_{kn}^* = \{u_k^*, \cdots, u_n^*\}$. $p_{1n}^*$是全过程的最优策略，简

称**最优策略**(optimal policy). 从初始状态 $x_1(=x_1^*)$出发，过程按照 $p_{1n}^*$和状态转移方程演变所经历的状态序列$\{x_1^*,x_2^*,\cdots,x_{n+1}^*\}$称**最优轨线**(optimal trajectory).

### 4.2.2 基本定理和基本方程

动态规划发展的早期阶段，从简单逻辑出发给出了所谓最优性原理，然后在最优策略存在的前提下导出基本方程，再由这个方程求解最优策略. 后来在动态规划的应用过程中发现，最优性原理不是对任何决策过程普遍成立，它与基本方程不是无条件等价，二者之间也不存在任何确定的蕴含关系. 基本方程在动态规划中起着更为本质的作用.

为了叙述的方便，在下面的定理中指标函数取各阶段指标之和的形式，即(4.3).

**定理 4.2.1** 对于初始状态 $x_1\in X_1$，策略 $p_{1n}^*=\{u_1^*,u_2^*,\cdots,u_n^*\}\in P_{1n}(x_1)$是最优策略的充要条件是对于任意的 $k$，$1<k\leqslant n$，有

$$V_{1n}(x_1,p_{1n}^*)=\underset{p_{1\,k-1}\in P_{1\,k-1}(x_1)}{\text{opt}}\Big[V_{1\,k-1}(x_1,p_{1\,k-1})+\underset{p_{kn}\in P_{kn}(x_k)}{\text{opt}}V_{kn}(x_k,p_{kn})\Big],$$

其中 $x_k$ 是由 $x_1,p_{1\,k-1}$和状态转移方程 $x_{j+1}=T_j(x_j,u_j)(j=1,2,\cdots,k-1)$所确定的第 $k$ 阶段的状态.

**定理 4.2.2** 若 $p_{1n}^*\in P_{1n}(x_1)$是最优策略，则对于任意的 $k$，$1<k<n$，它的子策略 $p_{kn}^*$对于由 $x_1$ 和 $p_{1\,k-1}^*$确定的以 $x_k^*$ 为起点的第 $k$ 到 $n$ 后部子过程而言，也是最优策略.

这就是最优性原理，通常略述为：不论过去的状态和决策如何，对于前面的决策形成的当前的状态而言，余下的各个决策必定构成最优策略.

最优性原理给出的是最优策略的必要条件，它是定理 4.2.1

的推论.

**定理 4.2.3** $\{f_k(x_k)\}$, $\{u_k^*\}$ 分别是最优值函数序列和最优决策序列的充要条件是满足下列递推方程,

$$f_k(x_k) = \underset{u_k \in U_k(x_k)}{\text{opt}} [v_k(x_k, u_k) + f_{k+1}(x_{k+1})], k = n, n-1, \cdots, 2, 1.$$

或者表为

$$f_k(x_k) = v_k(x_k, u_k^*) + f_{k+1}(T_k(x_k, u_k^*)).$$

以及

$$f_{n+1}(x_{n+1}) = \varphi(x_{n+1}) \quad (\varphi \text{为已知函数}),$$

$$x_{k+1} = T_k(x_k, u_k), \quad x_k \in X_k.$$

这就是动态规划的基本方程,为解决实际问题提供了有效的计算方法.

定理 4.2.3 中的 $f_{n+1}(x_{n+1}) = \varphi(x_{n+1})$ 是决策过程的终端条件. 当 $x_{n+1}$ 只取固定的状态时称固定终端;当 $x_{n+1}$ 可在终端集合 $X_{n+1}$ 中变动时称自由终端.

### 4.2.3 后向算法和前向算法

#### 1 后向算法

定理 4.2.3 给出的基本方程是动态规划的**后向算法**(backward algorithm). 计算步骤是,利用终端条件从 $k=n$ 开始由后向前逆推基本方程,求得各阶段的最优决策和最优值函数,最后算出 $f_1(x_1)$ 时就得到了最优决策序列 $\{u_k^*(x_k), x_k \in X_k, k=1,2,\cdots,n\}$. 再按照状态转移方程 $x_{k+1}^* = T_k(x_k, u_k^*(x_k))$, 从 $k=1$ 开始由前向后确定 $x_k^*$, 序列 $\{x_k^*, k=1,2,\cdots,n\}$ 为最优轨线, $\{u_k^*(x_k^*), k=1, 2,\cdots,n\}$ 为最优策略.

#### 2 前向算法

当决策过程可逆时也可以用动态规划的**前向算法**(forward algorithm)求解. 设阶段变量 $k$ 和状态 $x_k$ 的定义不变,决策 $u_k$ 应

使得状态转移能由 $x_{k+1}$ 和 $u_k$ 确定 $x_k$，即有

$$x_k = \tilde{T}_k(x_{k+1}, u_k), k = 1,2,\cdots,n. \tag{4.7}$$

允许决策集合相应地改变为 $\tilde{U}_k(x_{k+1})$. 前向算法的基本方程为

$$f_{k+1}(x_{k+1}) = \underset{u_k \in \tilde{U}_k(x_{k+1})}{\text{opt}} [v_k(x_k, u_k) + f_k(x_k)], k = 1,2,\cdots,n. \tag{4.8}$$

始端条件是

$$f_1(x_1) = \varphi(x_1). \tag{4.9}$$

前向算法的计算过程和后向算法正好相反.

**3　后向算法的计算框图**

以自由终端、固定始端、指标函数取和的形式(4.3)的后向算法为例给出计算框图，其它情况容易在这个基础上修改得到.

如果状态 $x_k$ 和决策 $u_k$ 是连续变量，用数值方法求解时需按照精度要求进行离散化. 设状态 $x_k$ 的允许集合为

$$X_k = \{x_{ki} | i = 1,2,\cdots,n_k\}, k = 1,2,\cdots,n.$$

决策 $u_{ki}(x_{ki})$ 的允许集合为

$$U_{ki} = \{u_{ki}^{(j)} | j = 1,2,\cdots,n_{ki}\}, i = 1,2,\cdots,n_k, k = 1,2,\cdots,n.$$

状态转移方程和阶段指标应对 $x_k$ 的每个取值 $x_{ki}$ 和 $u_{ki}$ 的每个取值 $u_{ki}^{(j)}$ 计算，即 $T_k = T_k(x_{ki}, u_{ki}^{(j)})$，$v_k = v_k(x_{ki}, u_{ki}^{(j)})$. 最优值函数应对 $x_k$ 的每个取值 $x_{ki}$ 计算. 基本方程可以表为

$$f_k^{(j)}(x_{ki}) = v_k(x_{ki}, u_{ki}^{(j)}) + f_{k+1}(T_k(x_{ki}, u_{ki}^{(j)})),$$
$$f_k(x_{ki}) = \underset{j}{\text{opt}} f_k^{(j)}(x_{ki}),$$
$$j = 1,2,\cdots,n_{ki},\ i = 1,2,\cdots,n_k,\ k = n,\cdots,2,1. \tag{4.10}$$

一般化的自由终端条件为

$$f_{n+1}(x_{n+1\,i}) = \varphi(x_{n+1\,i}), i = 1,2,\cdots,n_{n+1}. \tag{4.11}$$

其中 $\varphi$ 为已知. 固定始端条件可表示为 $X_1 = \{x_1\} = \{x_1^*\}$.

按照(4.11)，(4.10)后向计算出 $f_1(x_1^*)$，为全过程最优值. 记状态 $x_{ki}$ 的最优决策为 $u_{ki}^*(x_{ki})$，由 $x_1^*$ 和 $u_{ki}^*(x_{ki})$ 按照状态转移方

程计算出最优状态，记作 $x_k^*$. 并得到对应的最优决策，记作 $u_k^*(x_k^*)$. 于是最优策略为$\{u_1^*(x_1^*),u_2^*(x_2^*),\cdots,u_n^*(x_n^*)\}$.

算法程序的框图如下.

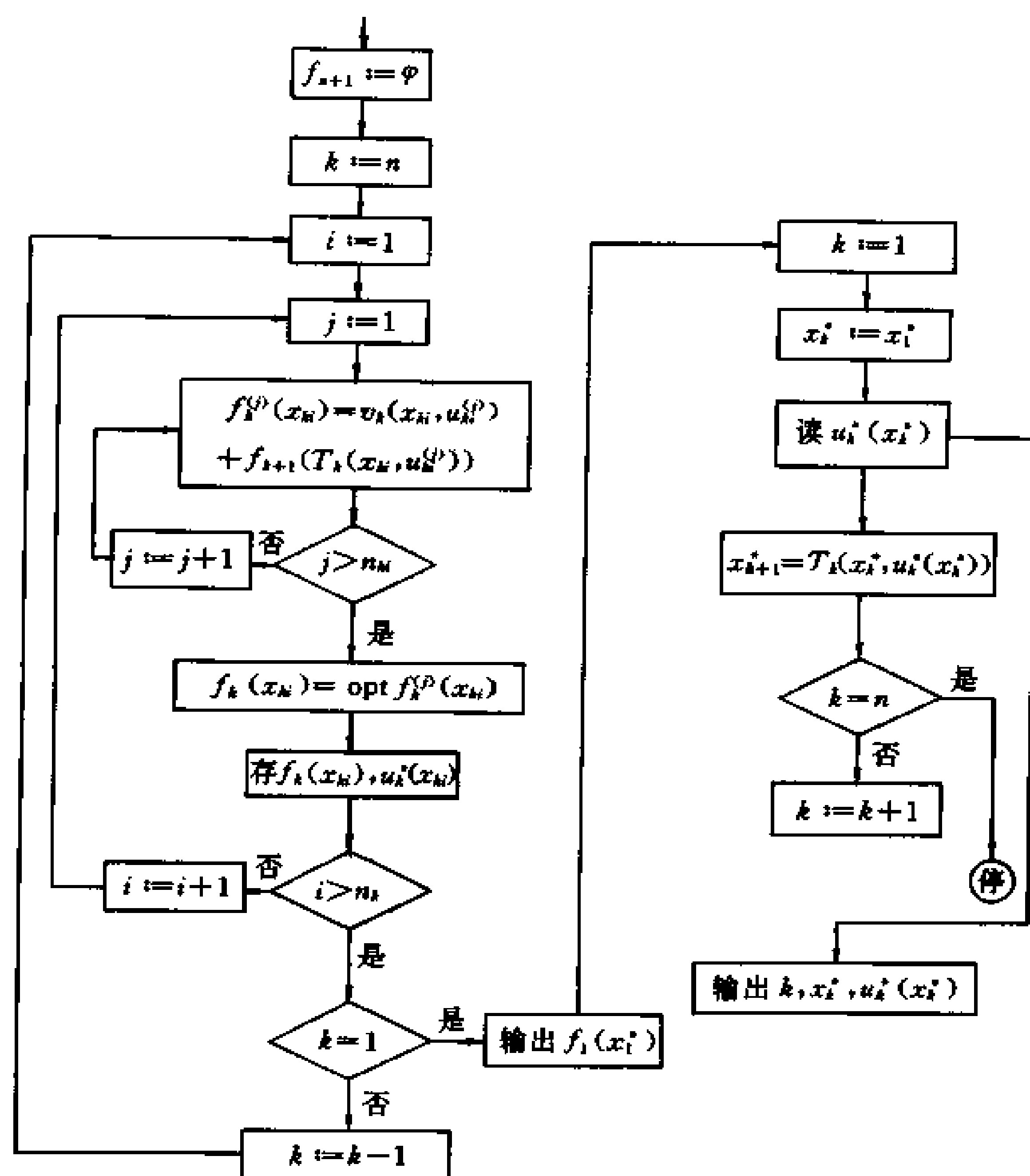



图 4.2

图的左边部分是函数序列的递推计算，可输出全过程最优值 $f_1(x_1^*)$，如果需要还可以输出后部子过程最优值函数序列 $f_k(x_{ki})$ 和最优决策序列 $u_k^*(x_{ki})$. 计算过程中存 $f_k(x_{ki})$ 是备计算 $f_{k-1}$ 之用，在 $f_{k-1}$ 算完后可用 $f_{k-1}$ 将 $f_k$ 替换掉；存 $u_k^*(x_{ki})$ 是备右边部分读 $u_k^*(x_k^*)$ 之用.

图的右边部分是最优状态和最优决策序列的前向计算，可输出最优策略 $\{u_1^*(x_1^*), u_2^*(x_2^*), \cdots, u_n^*(x_n^*)\}$ 和最优轨线 $\{x_1^*, x_2^*, \cdots, x_n^*\}$.

### 4.2.4 动态规划与静态规划的关系

动态规划与静态规划(线性和非线性规划等)研究的对象本质上都是在若干约束条件下的函数极值问题. 两种规划在很多情况下原则上可以相互转换.

动态规划可以看作求决策 $u_1, u_2, \cdots, u_n$ 使指标函数 $V_{1n}(x_1, u_1, u_2, \cdots, u_n)$ 达到最优(最大或最小)的极值问题，状态转移方程、端点条件以及允许状态集、允许决策集等是约束条件，原则上可以用非线性规划方法求解.

一些静态规划只要适当引入阶段变量、状态、决策等就可以用动态规划方法求解. 下面用例子说明.

**例 4.2.4** 用动态规划解下列非线性规划

$$\max \quad \sum_{k=1}^{n} g_k(u_k);$$

$$\text{s.t.} \quad \sum_{k=1}^{n} u_k = a, u_k \geqslant 0.$$

其中 $g_k(u_k)$ 为任意的已知函数.

**解** 按变量 $u_k$ 的序号划分阶段，看作 $n$ 段决策过程. 设状态为 $x_1, x_2, \cdots, x_n$，取问题中的变量 $u_1, u_2, \cdots, u_n$ 为决策. 状态转移方程为

$$x_1 = a,\ x_{k+1} = x_k - u_k,\ k = 1,2,\cdots,n.$$

取 $g_k(u_k)$为阶段指标，最优值函数的基本方程为(注意到 $x_{n+1}=0$)

$$f_k(x_k) = \max_{0\leqslant u_k\leqslant x_k}[g_k(u_k) + f_{k+1}(x_{k+1})];$$

$$0 \leqslant x_k \leqslant a, k = n, n-1, \cdots, 2, 1;$$

$$f_{n+1}(0) = 0.$$

按照后向算法求出对应于 $x_k$ 每个取值的最优决策 $u_k^*(x_k)$，计算至 $f_1(a)$后即可利用状态转移方程得到最优状态序列$\{x_k^*\}$和最优决策序列$\{u_k^*(x_k^*)\}$.

与静态规划相比，动态规划的优越性在于：

(1) 能够得到全局最优解. 由于约束条件确定的约束集合往往很复杂，即使指标函数较简单，用非线性规划方法也很难求出全局最优解. 而动态规划方法把全过程化为一系列结构相似的子问题，每个子问题的变量个数大大减少，约束集合也简单得多，易于得到全局最优解. 特别是对于约束集合、状态转移和指标函数不能用分析形式给出的优化问题，可以对每个子过程用枚举法求解，而约束条件越多，决策的搜索范围越小，求解也越容易. 对于这类问题，动态规划通常是求全局最优解的唯一方法.

(2) 可以得到一族最优解. 与非线性规划只能得到全过程的一个最优解不同，动态规划得到的是全过程及所有后部子过程的各个状态的一族最优解. 有些实际问题需要这样的解族，即使不需要，它们在分析最优策略和最优值对于状态的稳定性时也是很有用的. 当最优策略由于某些原因不能实现时，这样的解族可以用来寻找次优策略.

(3) 能够利用经验提高求解效率. 如果实际问题本身就是动态的，由于动态规划方法反映了过程逐段演变的前后联系和动态特征，在计算中可以利用实际知识和经验提高求解效率. 如在策略迭代法中，实际经验能够帮助选择较好的初始策略，提高收敛

速度.

动态规划的主要缺点是：

(1) 没有统一的标准模型，也没有构造模型的通用方法，甚至还没有判断一个问题能否构造动态规划模型的准则. 这样就只能对每类问题进行具体分析，构造具体的模型. 对于较复杂的问题在选择状态、决策、确定状态转移规律等方面需要丰富的想象力和灵活的技巧性，这就带来了应用上的局限性.

(2)用数值方法求解时存在**维数灾**(curse of dimensionality). 若一维状态变量有 $m$ 个取值，那么对于 $n$ 维问题，状态 $x_k$ 就有 $m^n$ 个值，对于每个状态值都要计算、存储函数$f_k(x_k)$，对于 $n$ 稍大(即使 $n=3$)的实际问题的计算往往是不现实的. 目前还没有克服维数灾的有效的一般方法.

## 4.3 若干典型问题的动态规划模型

建立一个确定性多阶段决策过程的动态规划模型包括以下几个步骤：

(1) 将过程划分成恰当的阶段.

(2) 正确选择状态变量 $x_k$，使它既能描述过程的状态，又满足无后效性，同时确定允许状态集合 $X_k$.

(3) 选择决策变量 $u_k$，确定允许决策集合 $U_k(x_k)$.

(4) 写出状态转移方程.

(5) 确定阶段指标 $v_k(x_k, u_k)$ 及指标函数 $V_{kn}$ 的形式(阶段指标之和，阶段指标之积，阶段指标之极大或极小等).

(6) 写出基本方程即最优值函数满足的递推方程，以及端点条件.

下面给出若干类典型问题的动态规划模型.

### 4.3.1 最短路线问题

对于例 4.1.1 一类**最短路线问题**(shortest path problem),阶段按过程的演变划分,状态由各段的位置确定,决策为从各个状态出发的走向,即有 $x_{k+1}=u_k(x_k)$,阶段指标为相邻两段状态间的距离 $d_k(x_k,u_k(x_k))$,指标函数为阶段指标之和,最优值函数 $f_k(x_k)$ 是由 $x_k$ 出发到终点的最短距离(或最小费用),基本方程为

$$f_k(x_k)=\min_{u_k(x_k)}[d_k(x_k,u_k(x_k))+f_{k+1}(x_{k+1})],k=n,\cdots,1;$$

$$f_{n+1}(x_{n+1})=0.$$

利用这个模型可以算出例 4.1.1 的最短路线为

$$AB_1C_2D_1E_2F_2G,$$

相应的最短距离为 18.

### 4.3.2 生产计划问题

对于例 4.1.2 一类**生产计划问题**(production planning problem),阶段按计划时间自然划分,状态定义为每阶段开始时的储存量 $x_k$,决策为每个阶段的产量 $u_k$,记每个阶段的需求量(已知量)为 $d_k$,则状态转移方程为

$$x_{k+1}=x_k+u_k-d_k,x_k\geqslant 0,k=1,2,\cdots,n. \tag{4.12}$$

设每阶段开工的固定成本费为 $a$,生产单位数量产品的成本费为 $b$,每阶段单位数量产品的储存费为 $c$,阶段指标为阶段的生产成本和储存费之和,即

$$v_k(x_k,u_k)=cx_k+\begin{cases}a+bu_k, u_k>0,\\0, \quad\quad u_k=0.\end{cases} \tag{4.13}$$

指标函数 $V_{kn}$ 为 $v_k$ 之和. 最优值函数 $f_k(x_k)$ 为从第 $k$ 段的状态 $x_k$ 出发到过程终结的最小费用,满足

$$f_k(x_k)=\min_{u_k\in U_k}[v_k(x_k,u_k)+f_{k+1}(x_{k+1})],k=n,\cdots,1. \tag{4.14}$$

其中允许决策集合 $U_k$ 由每阶段的最大生产能力决定. 若设过程终结时允许储存量为 $x_{n+1}^0$，则终端条件是

$$f_{n+1}(x_{n+1}^0) = 0. \tag{4.15}$$

(4.12)～(4.15)构成该问题的动态规划模型.

### 4.3.3 货物存储问题

考察确定性需求下，按周期订货，订货立即到达，使总费用(订货费、储存费和缺货损失费)达到最小的多阶段决策过程，讨论这类**货物存储问题**(inventory problem)的动态规划模型.

**例 4.3.1** 公司每周初订货，订货固定费 $a$，单位数量订货费 $b$，订货即刻进货. 每周需求量已知，记第 $k$ 周为 $w_k$. 如周末存储量为正值，需付存储费，单位数量存储费 $c_1$；如周末存储量为负值(即需求超过存储)，需付缺货损失费，单位数量损失费 $c_2$，且下周需补足缺货量. 仓库最大存储量为 $m_1$，允许最大缺货量 $m_2$. 试制订一个 $n$ 周的订货计划使总费用最小. $n$ 周结束后的库存货物可以变卖，变卖价值为已知量.

**解** 每周为一阶段，阶段变量 $k=1,2,\cdots,n$. 周初存储量为状态 $x_k$，订货量为决策 $u_k$，在需求 $w_k$ 下状态转移方程为

$$x_{k+1} = x_k + u_k - w_k,\ k = 1,2,\cdots,n. \tag{4.16}$$

订货费为

$$g_k(u_k) = \begin{cases} a + bu_k, u_k > 0, \\ 0, \quad\quad u_k = 0. \end{cases} \tag{4.17}$$

存储或损失费为

$$\begin{aligned} h_k(x_k,u_k) =& c_1\max(x_k + u_k - w_k,0) \\ &+ c_2\max(w_k - x_k - u_k,0). \end{aligned} \tag{4.18}$$

阶段指标为

$$v_k(x_k,u_k) = g_k(u_k) + h_k(x_k,u_k). \tag{4.19}$$

指标函数是阶段指标之和. 注意到最大存储量 $m_1$ 和最大缺货量

$m_2$ 的限制，最优值函数 $f_k(x_k)$ 的基本方程为

$$f_k(x_k)=\min_{0\leqslant u_k\leqslant m_1-x_k}[v_k(x_k,u_k)+f_{k+1}(x_{k+1})],$$
$$-m_2\leqslant x_k\leqslant m_1,\ k=n,n-1,\cdots,2,1. \qquad (4.20)$$

终端条件可表为

$$f_{n+1}(x_{n+1})=\varphi(x_{n+1}).$$

若进一步假设 $n$ 周后库存货物单位数量的价值为 $\beta$，且不许缺货，则有

$$\varphi(x_{n+1})=-\beta x_{n+1},\ 0\leqslant x_{n+1}\leqslant m_1. \qquad (4.21)$$

(4.16)～(4.21)即为本问题的动态规划模型.

### 4.3.4 设备更新问题

一台设备或一辆运输车随着使用期限的增加，由于效率降低收入减少，而维修费用却在增长. 使用多长时间进行更新可使某一时间内的总收入达到最大(或总费用达到最小)，下面两个例子说明这类**设备更新问题**(equipment replacing problem)的动态规划模型.

**例 4.3.2** 一台设备每年初要作出"继续使用"还是"更新"的决定. 假设当设备年龄为 $t$ 时，一年的收入为 $r(t)$，一年的维修费为 $d(t)$，更新的净费用为 $q(t)$(新设备的购价扣除旧设备的回收). 已知 $k=1$ 年初有一台年龄为 $s$ 的设备，试制订一个 $n$ 年的更新计划，使总收入最大. 设第 $n$ 年末年龄为 $t$ 的设备的回收 $\varphi(t)$ 为已知.

**解** 状态 $x_k$ 为第 $k$ 年设备的年龄 $t$，决策 $u_k$ 为第 $k$ 年更新设备，记作 $R$(replace)，或继续使用，记作 $K$(keep). 于是状态转移为

$$x_{k+1}=\begin{cases}1, & u_k=R,\\ t+1, & u_k=K.\end{cases}$$

阶段指标为一年的收入，即

$$v_k(t,u_k)=\begin{cases}r(0)-d(0)-q(t),\ u_k=R,\\ r(t)-d(t), \qquad u_k=K.\end{cases}$$

指标函数为阶段指标之和，最优值函数 $f_k(t)$ 为第 $k$ 年初一台年龄为 $t$ 的设备到过程终结可得到的最大收入，满足

$$f_k(t)=\max\begin{bmatrix}r(0)-d(0)-q(t)+f_{k+1}(1),\ u_k=R\\ r(t)-d(t)+f_{k+1}(t+1),\ u_k=K\end{bmatrix},$$

$$0\leqslant t\leqslant s+n,\ k=n,n-1,\cdots,2,1. \tag{4.22}$$

终端条件是

$$f_{n+1}(t)=\varphi(t). \tag{4.23}$$

由(4.23)、(4.22)递推算至 $f_1(s)$ 即为全过程最大收入，并得到最优策略.

如果收入、维修费等除与设备年龄 $t$ 有关外，还与时间有关，并且要考虑折扣因子 $\alpha(0\leqslant\alpha\leqslant1)$，表示一年以后收入的实际价值仅相当于现年的 $\alpha$ 倍，那么(4.22)应修改为

$$f_k(t)=\max\begin{bmatrix}r_k(0)-d_k(0)-q_k(t)+\alpha f_{k+1}(1),\ u_k=R\\ r_k(t)-d_k(t)+\alpha f_{k+1}(t+1),\ u_k=K\end{bmatrix}.$$

**例 4.3.3** 在例 4.3.2 的基础上增加一个可以采取的决策——修复，即一次全面整修，能使旧设备的收入增加，维修费降低，修复费用又低于更新的费用. 这样原来定义的收入和费用除与设备年龄 $t$ 有关外，还取决于它在“几岁”时修复过. 于是已知条件是：年龄 $t$ 且在 $s$ 岁修复过的设备年收入为 $r(t,s)$；年维修费为 $d(t,s)$；更新费是 $q(t,s)$. $n$ 年末的回收为 $\varphi(t,s)$. 以上 $s$ 指最后一次修复时的年龄. 又年龄 $t$ 时的修复费为 $h(t)$. 试重新构造使总收入最大的动态规划模型.

**解** 状态 $x_k$ 应定义为二维变量 $(t,s)$，$t$ 是第 $k$ 年设备的年龄，$s$ 是它最后一次修复时的年龄 $(s\leqslant t)$. 增加的决策“修复”记作 $O$(overhaul). 最优值函数 $f_k(t,s)$ 满足的基本方程为（不妨设 $k=1$ 时是一台新设备）

$$f_k(t,s)=\max\begin{bmatrix} r(0,0)-d(0,0)-q(t,s)+f_{k+1}(1,0), & u_k=R \\ r(t,s)-d(t,s)+f_{k+1}(t+1,s), & u_k=K \\ r(t,t)-d(t,t)-h(t)+f_{k+1}(t+1,t), & u_k=O \end{bmatrix},$$

$$0\leqslant t\leqslant n,\ s\leqslant t,\ k=n,\cdots,2,1. \tag{4.24}$$

$$f_{n+1}(t,s)=\varphi(t,s). \tag{4.25}$$

由(4.25),(4.24)递推计算至 $f_1(0,0)$ 即为全过程最大收入,并得到最优策略.

### 4.3.5 资源分配问题

一种或几种资源(包括资金)分配给若干用户,或投资于几家企业,以获得最大的效益. **资源分配问题**(resource allocating problem)可以是多阶段决策过程,也可以是静态规划问题,都能构造动态规划模型求解. 下面分别举例说明.

**例 4.3.4** 机器可以在高、低两种负荷下生产. $u$ 台机器在高负荷下的年产量是 $g(u)$,在低负荷下的年产量是 $h(u)$,高、低负荷下机器的年损耗率分别是 $a_1$ 和 $b_1$ $(0<b_1<a_1<1)$. 现有 $m$ 台机器,要安排一个 $n$ 年的负荷分配计划,即每年初决定多少台机器投入高、低负荷运行,使 $n$ 年的总产量最大. 如果进一步假设 $g(u)=\alpha u$, $h(u)=\beta u$ $(\alpha>\beta>0)$,即高、低负荷下每台机器的年产量分别为 $\alpha$ 和 $\beta$,结果将有什么特点.

**解** 年度为阶段变量 $k=1,2,\cdots,n$. 状态 $x_k$ 为第 $k$ 年初完好的机器数,决策 $u_k$ 为第 $k$ 年投入高负荷运行的台数. 当 $x_k$ 或 $u_k$ 不是整数时,将小数部分理解为一年中正常工作时间或投入高负荷运行时间的比例.

机器在高、低负荷下的年完好率分别记为 $a$ 和 $b$,则 $a=1-a_1$, $b=1-b_1$,有 $a<b$. 因为第 $k$ 年投入低负荷运行的机器台数为 $x_k-u_k$,所以状态转移方程是

$$x_{k+1} = au_k + b(x_k - u_k). \tag{4.26}$$

阶段指数 $v_k$ 是第 $k$ 年的产量，有

$$v_k(x_k, u_k) = g(u_k) + h(x_k - u_k). \tag{4.27}$$

指标函数是阶段指标之和，最优值函数 $f_k(x_k)$ 满足

$$f_k(x_k) = \max_{0 \leqslant u_k \leqslant x_k} [v_k(x_k, u_k) + f_{k+1}(x_{k+1})],$$
$$0 \leqslant x_k \leqslant m,\ k = n, \cdots, 2, 1. \tag{4.28}$$

及自由终端条件

$$f_{n+1}(x_{n+1}) = 0,\ 0 \leqslant x_{n+1} \leqslant m. \tag{4.29}$$

当 $v_k$ 中的 $g, h$ 用较简单的函数表达式给出时，对于每个 $k$ 可以用解析方法求解极值问题. 特别，若 $g(u) = \alpha u, h(u) = \beta u$，(4.28)中的$[v_k(x_k, u_k) + f_{k+1}(x_{k+1})]$将是 $u_k$ 的线性函数，最大值点必在区间 $0 \leqslant u_k \leqslant x_k$ 的左端点 $u_k = 0$ 或右端点 $u_k = x_k$ 取得，即每年初将完好的机器全部投入低负荷或高负荷运行.

**例 4.3.5** 总量为 $m_1$ 的资源 $A$ 和总量为 $m_2$ 的资源 $B$ 同时分配给 $n$ 个用户. 已知第 $k$ 用户利用数量 $u_k$ 的 $A$ 和数量 $v_k$ 的 $B$ 可以产生的效益为 $g_k(u_k, v_k)$. 问如何分配这些资源可使总效益最大. 试建立这个静态规划问题的动态规划模型.

**解** 这个典型的静态规划问题为

$$\max_{u_k, v_k} \sum_{k=1}^{n} g_k(u_k, v_k); \tag{4.30}$$

$$\text{s.t.}\quad \sum_{k=1}^{n} u_k = m_1,\ u_k \geqslant 0,$$
$$\sum_{k=1}^{n} v_k = m_2,\ v_k \geqslant 0. \tag{4.31}$$

当 $g_k$ 比较复杂或 $n$ 较大时，用非线性规划求解是困难的. 特别，若 $g_k$ 用表格或图形给出而没有解析表达式时则无法求解.

将(4.30)、(4.31)化为动态规划问题. 每分配给一个用户作为

一个阶段，记作 $k=1,2,\cdots,n$. 分配给第 $k$ 用户的资源数量是二维决策变量 $(u_k,v_k)$，而把向第 $k$ 用户分配时，分配者手中掌握的资源总量作为二维状态变量 $(x_k,y_k)$，状态转移方程为

$$\begin{cases} x_{k+1}=x_k-u_k, \\ y_{k+1}=y_k-v_k. \end{cases} \tag{4.32}$$

阶段指标为 $g_k(u_k,v_k)$，指标函数是 $g_k$ 之和，最优值函数 $f_k(x_k,y_k)$ 是将资源总量 $x_k,y_k$ 分给第 $k$ 至第 $n$ 用户能获得的最大效益，满足

$$f_k(x_k,y_k)=\max_{\substack{0\leqslant u_k\leqslant x_k \\ 0\leqslant v_k\leqslant y_k}}[g_k(u_k,v_k)+f_{k+1}(x_{k+1},y_{k+1})];$$

$$0\leqslant x_k\leqslant m_1,0\leqslant y_k\leqslant m_2,k=n,\cdots,2,1. \tag{4.33}$$

终端条件为

$$f_{n+1}(0,0)=0. \tag{4.34}$$

(4.32)～(4.34)组成的动态规划模型与非线性规划(4.30)、(4.31)等价.

### 4.3.6 系统可靠性问题

一个系统由若干部件串接组成，只要有一个部件出现故障，整个系统就不能正常工作. 为提高系统的可靠性，每个部件都装有备用件，一旦原部件出现故障，备用件就自动进入系统. 显然，备用件越多系统可靠性越大，但费用也越高. 所谓**系统可靠性问题**(system reliability problem)，即在一定的总费用下如何配置各部件的备件使系统可靠性最大，也是用动态规划求解的一类典型问题.

**例 4.3.6** 由 $n$ 个部件串接而成的系统，当部件 $k$ 配置 $u_k$ 个备用件时，这个部件正常工作的概率为 $p_k(u_k)$，而每个备用件的费用是 $c_k$. 试在总费用不超过 $c$ 的条件下决定各个部件备用件的数量，使得系统正常工作的概率最大.

**解** 因为系统正常工作的概率是各部件正常工作概率的乘

积，所以问题归结为

$$\max_{u_k} \prod_{k=1}^{n} p_n(u_k); \tag{4.35}$$

$$\text{s.t.} \quad \sum_{k=1}^{n} c_k u_k \leqslant c, \ u_k \text{是正整数或零}.$$

直接求解这个非线性整数规划比较困难，可以转化为动态规划问题.

按照对 $n$ 个部件配置备用件的次序划分阶段 $k=1,2,\cdots,n$. 部件 $k$ 配置备用件的数量为决策 $u_k$，根据容许费用的限制，为部件 $k$ 配置备件时所容许的费用为状态 $x_k$，于是状态转移方程为

$$x_{k+1} = x_k - c_k u_k. \tag{4.36}$$

阶段指标 $v_k$ 是部件 $k$ 正常工作的概率 $p_k(u_k)$，而指标函数 $V_{kn}$ 应是阶段指标的乘积形式，即

$$V_{kn} = \prod_{j=k}^{n} p_j(u_j).$$

最优值函数 $f_k(x_k)$ 定义为状态 $x_k$ 下由部件 $k$ 到 $n$ 组成的子系统的最大正常工作概率，满足

$$f_k(x_k) = \max_{u_k \in U_k(x_k)} [p_k(u_k) f_{k+1}(x_{k+1})],$$

$$U_k(x_k) = \left\{ u_k \middle| 0 \leqslant u_k \leqslant \frac{x_k}{c_k}, \text{且为整数} \right\},$$

$$0 \leqslant x_k \leqslant c, \ k = n, \cdots, 2, 1. \tag{4.37}$$

终端条件为

$$f_{n+1}(x_{n+1}) = 1. \tag{4.38}$$

这是因为在(4.37)中 $k=n$ 时应有

$$f_n(x_n) = \max_{u_n \in U_n(x_n)} p_n(u_n).$$

利用(4.36)～(4.38)计算出 $f_1(c)$，即为系统最大正常工作概率，同时可得最优策略 $\{u_1^*, \cdots, u_n^*\}$.

### 4.3.7 任务均衡问题

一批任务由若干设备或机构完成，每一个设备或机构完成一项任务的时间是不同的，如何向各个设备均衡地分配任务，使这批任务能尽快地完成. 下面举例说明这类**任务均衡问题**(job balancing problem)的动态规划模型.

**例 4.3.7** $n$ 台不同的车床同时加工 $m$ 个相同的零件，将车床由 1 到 $n$ 编号，已知在车床 $k$ 上加工一个零件的时间为 $t_k(k=1,2,\cdots,n)$. 如何给 $n$ 台车床分配零件加工，使完工时间最短.

**解** 设向车床 $k$ 分配 $u_k$ 个零件，则车床 $k$ 加工时间为 $u_kt_k$. 因为整个任务的完工时间由加工时间最长的那台机床决定，所以完工时间为$\max\limits_k u_kt_k$. 于是这个问题的静态模型是

$$\min_{u_k} \quad (\max_{1\leqslant k\leqslant n} u_kt_k);$$

$$\text{s.t.} \quad \sum_{k=1}^{n} u_k = m, u_k \text{ 为非负整数.} \tag{4.39}$$

构造动态规划模型时以按顺序向各台车床分配任务划分阶段 $k=1,2,\cdots,n$. 决策 $u_k$ 即为车床 $k$ 分配的零件数，而状态 $x_k$ 定义为向车床 $k$ 分配任务时分配者手中尚存的零件数. 于是状态转移规律为

$$x_{k+1} = x_k - u_k. \tag{4.40}$$

阶段指标显然是 $v_k(x_k,u_k)=u_kt_k$. 指标函数则为

$$V_{kn}(x_k,u_k,\cdots,x_{k+1}) = \max_{k\leqslant j\leqslant m} v_j(x_j,u_j).$$

最优值函数 $f_k(x_k)$满足

$$f_k(x_k) = \min_{u_k\in U_k(x_k)} [\max(u_kt_k, f_{k+1}(x_{k+1}))],$$

$$U_k(x_k) = \{u_k | u_k = 0,1,\cdots,x_k\},$$

$$x_k = 0,1,\cdots,m,\ k = n, n-1,\cdots,2,1. \tag{4.41}$$

终端条件为

$$f_{n+1}(0) = 0. \tag{4.42}$$

利用(4.40)～(4.42)递推计算至 $f_1(m)$，得到最短完工时间，再求出最优策略$\{u_1^*,\cdots,u_n^*\}$.

### 4.3.8 排序问题

在几台机器上加工一批互不相同的零件，已知每个零件在每台机器上加工时间，要安排一个加工次序使加工完全部零件所需总时间最短. 当机器数目较大时这是一个非常复杂的组合优化问题，下面只给出 2 台机器情况下**排序问题**(sequencing problem)的动态规划模型.

**例 4.3.8** $n$ 个零件要在机器 $A$，$B$ 上加工，每个零件都必须先经过 $A$，后经过 $B$ 两道工序. 零件 $i$ 在 $A$，$B$ 加工所需工时分别记作 $a_i$，$b_i$($i=1,2,\cdots,n$). 试安排各零件的加工次序使加工完 $n$ 个零件所需总时间最短.

**解** 首先，$n$ 个零件在 $A$ 上的加工次序应该与在 $B$ 上的加工次序相同，因为否则，必然有零件在 $A$ 加工完毕后，要等待其他零件在 $B$ 上加工完才能在 $B$ 上加工，使 $B$ 的等待加工时间变长. 这样，$n$ 个零件在 $A$，$B$ 上只有一个加工次序. 并且，零件在 $A$ 上加工没有等待时间，为使总时间最短应尽量减少零件在 $B$ 上的等待时间.

以在 $A$ 上更换零件的时刻划分阶段 $k=1,2,\cdots,n$，在每一阶段用 $S$ 表示在 $A$ 上等待加工的零件集合. 设 $x$ 是 $S$ 之前在 $A$ 上加工的最后一个零件，用 $t$ 表示从 $x$ 在 $A$ 上加工完到 $x$ 在 $B$ 上加工完所需的时间. 在每个阶段 $k$ 用$(S,t)$表示过程所处的状态，其中集合 $S$ 中零件的个数 $|S|=n-k+1$.

记 $f(S,t)$为从状态$(S,t)$出发将 $S$ 中零件按最优次序全部加工完所需的时间；$f(S,t,i)$为从$(S,t)$出发先加工 $i$ 再将 $S$ 中其余零件按最优次序全部加工完所需时间；$f(S,t,i,j)$为从$(S,t)$出发

相继加工 $i,j$ 后再将 $S$ 中其余零件按最优次序全部加工完所需时间，则它们满足

$$\begin{cases} f(S,t,i) = a_i + f(S\backslash i, z_i(t)), \\ z_i(t) = \max(t - a_i, 0) + b_i. \end{cases} \tag{4.43}$$

$$\begin{cases} f(S,t,i,j) = a_i + a_j + f(S\backslash(i \cap j), z_{ij}(t)), \\ z_{ij}(t) = \max(z_i(t) - a_j, 0) + b_j. \end{cases} \tag{4.44}$$

由此可以导出，当

$$z_{ij}(t) \leqslant z_{ji}(t) \tag{4.45}$$

时必有

$$f(S,t,i,j) \leqslant f(S,t,j,i). \tag{4.46}$$

故(4.45)是对于任意的$(S,t)$零件 $i$ 应排在零件 $j$ 前面加工的条件.

利用 $z_{ij}(t)$在(4.44)，(4.43)中的定义，条件(4.45)可以表示为

$$\min(a_i,b_j) \leqslant \min(a_j,b_i). \tag{4.47}$$

这就是零件 $i$ 应排在零件 $j$ 前面的条件. 由此给出如下的最优排序规则：

(1) 写出零件的加工时间矩阵

$$R = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}.$$

(2) 找出 $R$ 中的最小数(如果最小数不止一个，可任选其一). 若最小数为 $a_i$，则零件 $i$ 排在加工次序的第一位；若最小数为 $b_j$，则零件 $j$ 排在末一位.

(3) 在 $R$ 中去掉 $a_i$ 或 $b_j$ 所在的列，重复(2)，但注意应接着排在第一位之后或末一位之前. 如此直至将所有零件排完.

### 4.3.9 推销商问题

**推销商问题**(traveling salesman problem)是组合优化中的一

个著名问题，可以用动态规划方法求解.

**例 4.3.9** 有 $n$ 个城市，用 $1,2,\cdots,n$ 表示，城 $i,j$ 之间的距离为 $d_{ij}$. 一个推销商从城 1 出发到其它每个城市去一次且只去一次，最后回到城 1. 怎样选择行走路线使总路程最短.

**解** 推销商要经过 $n-1$ 个城市，阶段变量记作 $k=1,2,\cdots,n-1$. 设在第 $k$ 段推销商到达城 $i$ 并且途中经过的城市集合为 $S$，状态 $x_k$ 为 $(S,i)$，决策 $u_k$ 为他前往的下一城市 $j$，阶段指标为 $d_{ij}$，最优值函数 $f_k(S,i)$ 为由城 1 出发经过 $k$ 个城市的集合 $S$ 到达城 $i$ 的最短距离，满足

$$
\begin{aligned}
&f_k(S,i) = \min_{j\in S}[d_{ij} + f_{k-1}(S\setminus j,j)],\\
&S \subset N_i = \{2,3,\cdots,i-1,i+1,\cdots,n\},|S| = k,\\
&i = 2,3,\cdots,n,k = 1,2,\cdots,n-1.
\end{aligned}
\tag{4.48}
$$

式中 $|S|$ 表示 $S$ 中城市的数量. (4.48)是动态规划前向算法的递推方程. 始端条件为

$$
f_0(\varnothing,i) = d_{1i},i = 2,3,\cdots,n. \tag{4.49}
$$

利用(4.49)，(4.48)计算出 $f_{n-1}(N,1)$，$N=\{2,3,\cdots,n\}$，即为全程的最短距离，同时可以得到最优策略，即最优行走路线.

### 4.3.10 线性系统的二次指标函数问题

线性系统中求控制函数使某个二次指标达到最大(小)，是最优控制问题的最基本的形式. 对于离散线性系统，这是一个多阶段决策过程，用动态规划方法求解十分方便. **二次指标函数问题**(quadratic objective function problem)的标准形式为，

$$
\min_{x_1,\cdots,x_n}\left[\sum_{k=1}^{n}(\boldsymbol{x}_k^{\mathrm T}\boldsymbol{Q}_k\boldsymbol{x}_k + 2\boldsymbol{x}_k^{\mathrm T}\boldsymbol{S}_k\boldsymbol{u}_k + \boldsymbol{u}_k^{\mathrm T}\boldsymbol{R}_k\boldsymbol{u}_k) + \boldsymbol{x}_{n+1}^{\mathrm T}\boldsymbol{Q}_{n+1}\boldsymbol{x}_{n+1}\right]; \tag{4.50}
$$

$$
\begin{aligned}
\text{s.t.}\quad &\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\boldsymbol{x}_k + \boldsymbol{B}_k\boldsymbol{u}_k,\ \boldsymbol{x}_k \in X_k,\\
&k = 1,2,\cdots,n.
\end{aligned}
\tag{4.51}
$$

其中 $\boldsymbol{x}_k$ 为 $p$ 维状态变量，$\boldsymbol{u}_k$ 为 $r$ 维控制变量，$\boldsymbol{A}_k$，$\boldsymbol{B}_k$ 为相应维数的矩阵，$\boldsymbol{R}_k$ 为正定阵，$\boldsymbol{Q}_k$ 为半正定阵. 对 $\boldsymbol{u}_k$ 没有限制.

以 $\boldsymbol{x}_k$ 为动态规划模型的状态，$\boldsymbol{u}_k$ 为决策，(4.51)为状态转移方程，阶段指标为

$$v_k(x_k,u_k) = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q}_k\boldsymbol{x}_k + 2\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{u}_k + \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{R}_k\boldsymbol{u}_k. \tag{4.52}$$

最优值函数 $f_k(x_k)$满足

$$f_k(\boldsymbol{x}_k) = \min_{u_k}[v_k(\boldsymbol{x}_k,\boldsymbol{u}_k) + f_{k+1}(\boldsymbol{x}_{k+1})], k = n,\cdots,2,1.$$

$$f_{n+1}(\boldsymbol{x}_{n+1}) = \boldsymbol{x}_{n+1}^{\mathrm{T}}\boldsymbol{Q}_{n+1}\boldsymbol{x}_{n+1}. \tag{4.53}$$

将(4.51)，(4.52)代入(4.53)可知，这是一个 $\boldsymbol{u}_k$ 的二次函数极值问题，可以用下述定理给出最优解的解析表达式.

**定理 4.3.10** 对于线性系统的二次指标函数的最优控制问题(4.50)，(4.51)，最优解为

$$\begin{aligned}
&\boldsymbol{u}_k^*(\boldsymbol{x}_k) = -\boldsymbol{L}_k\boldsymbol{x}_k, k = 1,2,\cdots,n;\\
&\boldsymbol{L}_k = (\boldsymbol{R}_k + \boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)^{-1}(\boldsymbol{S}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)^{\mathrm{T}};\\
&\boldsymbol{P}_k = \boldsymbol{Q}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{A}_k - (\boldsymbol{S}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)\boldsymbol{L}_k,\\
&\boldsymbol{P}_{n+1} = \boldsymbol{0}, \qquad k = 1,2,\cdots,n.
\end{aligned} \tag{4.54}$$

并且(4.53)的最优值函数为

$$f_k(\boldsymbol{x}_k) = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{P}_k\boldsymbol{x}_k, k = 1,2,\cdots,n. \tag{4.55}$$

在最优控制函数序列$\{\boldsymbol{u}_1^*,\cdots,\boldsymbol{u}_n^*\}$下，由 $x_1$ 出发的全过程最优指标是 $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{P}_1\boldsymbol{x}_1$，最优轨线$\{\boldsymbol{x}_1^*,\cdots,\boldsymbol{x}_n^*\}$由 $\boldsymbol{u}_k^*(\boldsymbol{x}_k)$和(4.51)确定.

以上假定 $\boldsymbol{u}_k^*(\boldsymbol{x}_k)\in U_k(\boldsymbol{x}_k)$，$\boldsymbol{x}_k^*\in X_k$.

定理 4.3.10 表明，最优控制 $\boldsymbol{u}_k^*(\boldsymbol{x}_k)$是状态 $\boldsymbol{x}_k$ 的线性负反馈.

## 4.4 不定期和无限期决策过程

在 4.2 和 4.3 中多阶段决策过程的阶段数 $n$ 固定，是定期决策过程. 当 $n$ 不固定时称为**不定期决策过程**(decision process of

unfixed step number). 当 $n$ 趋向无穷时称为**无限期决策过程**(decision process of infinite step number). 用动态规划方法求解不定期和无限期决策过程的优化问题时,用到的基本概念和原理与定期过程是相同的,只是基本方程的形式有所不同,因而方程的解法也不同. 下面先通过两个例子说明这两种过程的背景及其动态规划模型.

**例 4.4.1** 段数不定的最短路线问题

$n$ 个点相互连接组成一个线路网(右图中 $n=5$),各点标号为 $1,2,\cdots,n$. 任两点 $i,j$ 之间的距离(或费用)记作 $d_{ij}$(当 $i,j$ 不直接相连时令 $d_{ij}=\infty$). 求任意一点 $i(i=1,2,\cdots,n-1)$到点 $n$ 的最短路线.

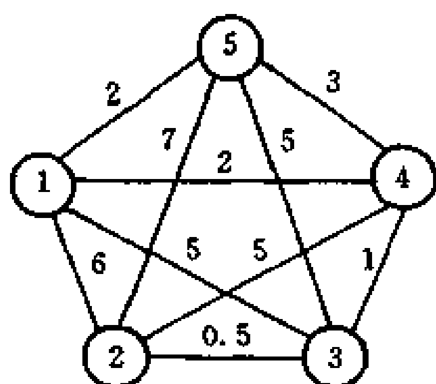


图 4.3

**解** 可以不考虑回路,因为含有回路的路线一定不是最短的.

问题中路线的段数事先不固定,而是随着最优策略确定的. 这里的状态、决策、状态转移、指标函数等都与 4.3.1 最短路线问题相同,但是它们都不再依赖于阶段变量 $k$,状态记作 $x=i,i=1,2,\cdots,n$,决策记作 $u(i)$. 策略是对于任意状态 $x$ 的决策函数,记作 $u(x)$. 阶段指标是任意两状态 $i,j$ 间的距离 $d_{ij}$,指标函数 $V(i,u(x))$是由状态 $i$ 出发,在策略 $u(x)$ 下到达状态 $n$ 的路线的距离,它是阶段指标之和,并满足可分离性要求,有

$$V(i,u(x)) = d_{ij} + V(j,u(x)).$$

最优值函数 $f(i)$为由 $i$ 出发到达 $n$ 的最短距离,即

$$f(i) = \min_{u(x)} V(i,u(x)) = V(i,u^*(x)),$$

式中 $u^*(x)$是最优策略,类似于定理 4.2.3. $f(i)$满足基本方程

$$f(i) = \min_{1 \leqslant j \leqslant n} [d_{ij} + f(j)], i = 1,2,\cdots,n-1. \quad (4.56)$$

这是一个关于 $f(i)$的函数方程，对固定的 $i$ 使(4.56)右端$[d_{ij}+f(j)]$达到极小的 $j$ 即为最优决策 $u^*(i)$. 对所有的 $i$ 求解(4.56)得到最优策略 $u^*(x)$.

**例 4.4.2** 无限期的负荷分配问题

在例 4.3.4 机器高、低负荷分配问题中，考察无限期即 $n \to \infty$ 的情况，建立相应的动态规划模型.

**解** 首先，从状态转移方程 $x_{k+1} = au_k + b(x_k - u_k)$和允许决策集合 $0 \leqslant u_k \leqslant x_k$ 可知，从初始状态 $x_1 = m$（开始时拥有的机器台数）出发经 $n$ 年后，不论这 $n$ 年的负荷分配方案如何，状态 $x_{n+1}$（$n$ 年后完好的机器台数）都满足

$$x_{n+1} \geqslant ma^n \qquad (0 < a < 1),$$

所以负荷分配可以无限期地进行下去.

其次，当阶段指标为 $v_k(x_k, u_k) = \alpha u_k + \beta(x_k - u_k)$时，从初始状态 $x_1 = m$ 转移至第 $k$ 年，不论其分配方案如何，第 $k$ 年的阶段指标 $v_k$ 都满足

$$v_k(x_k, u_k) \leqslant \alpha x_k \leqslant \alpha m b^{k-1}.$$

于是 $n$ 年的指标函数

$$V_{1n}(x_1, p_{1n}) \leqslant \alpha m \sum_{k=1}^{n} b^{k-1} \quad (0 < b < 1)$$

是单调增有界的，$\lim\limits_{n \to \infty} V_{1n}$存在. 可以将此极限值定义为无限期决策过程的指标函数. 相应地，最优值函数 $f_k(x_k)$也定义为定期过程最优值函数当 $n \to \infty$时的极限值，它满足与定期决策过程相同的基本方程.

对于无限期过程，不能用基本方程的递推形式进行计算，但是因为允许决策集合

$$U(x_k) = \{u_k | u_k = 0, 1, \cdots, x_k\},$$

状态转移方程

$$x_{k+1}=T(x_k,u_k)=au_k+b(x_k-u_k),$$

阶段指标

$$v(x_k,u_k)=\alpha u_k+\beta(x_k-u_k),$$

均与阶段变量 $k$ 无关，并且可将 $x_k$ 记作 $x$，$x\in X=\{x\mid 0<x\leqslant m\}$，故基本方程可以去掉所有的下标 $k$，写作

$$f(x)=\max_{u\in U(x)}[v(x,u)+f(T(x,u))], \tag{4.57}$$

其中

$$\begin{aligned} v(x,u)&=\alpha u+\beta(x-u),\\ T(x,u)&=au+b(x-u),\\ U(x)&=\{u\mid 0\leqslant u\leqslant x\}. \end{aligned} \tag{4.58}$$

(4.57)，(4.58)是 $f(x)$ 的函数方程，对于 $x\in X$ 求出最优解 $u^*(x)$，就是每个阶段在状态 $x$ 下的最优负荷分配方案.

函数方程可以用迭代法求解，通常有以下两种迭代法.

### 4.4.1 函数迭代法和策略迭代法

以例 4.4.1 的函数方程(4.56)为例说明这两种迭代方法. **函数迭代法**(function iterative method)的具体步骤是：

(1) 选初始函数 $f_1(i)$ 为由 $i$ 一步到 $n$ 的最短距离，即

$$f_1(i)=\begin{cases} d_{in}, & i=1,2,\cdots,n-1,\\ 0, & i=n. \end{cases} \tag{4.59}$$

(2) 按照迭代公式

$$f_k(i)=\begin{cases} \min\limits_{1\leqslant j\leqslant n}[d_{ij}+f_{k-1}(j)], & i=1,2,\cdots,n-1,\\ 0, & i=n. \end{cases} \tag{4.60}$$

求 $f_k(i)$，$k=2,3,\cdots$.

(3) 当计算至

$$f_{k+1}(i) = f_k(i),\ i = 1,2,\cdots,n$$

时停止. $f_k(i)=f(i)$满足原来的函数方程(4.56),$k$ 不会超过 $n-1$. 由此求得最优决策 $u^*(i)$.

这种迭代法的基本思想是,先求 1 段内由 $i$ 到 $n$ 的最短路线,再求 2,3,…段内由 $i$ 到 $n$ 的最短路线. 因为共 $n$ 个点,所以全局的最短路线必在 $n-1$ 段内产生,否则将出现回路. 可以表述为如下定理.

**定理 4.4.3** 由函数迭代法确定的函数序列$\{f_k(i)\}$单调非增地收敛于 $f(i)$,$f(i)$满足基本方程 $f(i)=\min\limits_{1\leqslant j\leqslant n}[d_{ij}+f(j)]$,且迭代步数 $k$ 不超过 $n-1$.

**策略迭代法**(policy iterative method)的步骤是:

(1) 选一条没有回路的初始策略 $u_1(i)$,$i=1,2,\cdots,n-1$,表示由 $i$ 出发到达的下一点 $j$,$j=u_1(i)$.

(2) 令 $k=1$ 解代数方程组

$$f_k(i) = \begin{cases} d_{i,u_k(i)} + f_k(u_k(i)), & i = 1,2,\cdots,n-1, \\ 0, & i = n. \end{cases} \tag{4.61}$$

求出 $f_k(i)$.

(3) 由 $f_k(i)$计算策略

$$u_{k+1}(i) = (j\,|\min_{1\leqslant j\leqslant n}[d_{ij} + f_k(j)]),\ i = 1,2,\cdots,n-1. \tag{4.62}$$

(4) 若

$$u_{k+1}(i) = u_k(i),\ i = 1,2,\cdots,n-1,$$

则迭代停止,最优策略 $u^*(i)=u_k(i)$;否则以 $k+1$ 代替 $k$ 重复(2),(3).

这种方法的基本思想是对于初始的及每步得到的策略求解函数方程,然后按照优化的方向改进策略,直至收敛. 这时函数序列

$\{f_k(i)\}$也收敛于(4.56)的 $f(i)$，有以下定理.

**定理 4.4.4** 若初始策略 $u_1(i)$没有回路，则由策略迭代法确定的$\{u_k(i)\}$也没有回路，且$\{f_k(i)\}$单调非增地收敛于 $f(i)$，$f(i)$满足基本方程 $f(i)=\min\limits_{1\leqslant j\leqslant n}[d_{ij}+f(j)]$.

在策略迭代法的每次迭代中有求值 $f_k(x)$和改进策略 $u_k(x)$两部分.求值时要解一个未知数个数等于状态取值数目的代数方程组，当状态取值数目较大时计算量很大，所以就每次迭代而言策略迭代法比函数迭代法复杂.但是它所需的迭代次数往往比较少，特别是对有一定知识和经验的实际问题，可以选取一个较好的初始策略，收敛会很快.相比之下，用知识和经验选择初始函数则比较困难.

### 4.4.2 不定期平稳决策过程和平稳策略

例 4.4.1 和 4.4.2 的共同特点是在多阶段决策过程中允许决策集合、状态转移规律、阶段指标等与阶段变量 $k$ 无关，从而基本方程成为函数方程，称这样的过程是**平稳的**.

**定义 4.4.5** 满足以下条件的多阶段决策过程称为**平稳过程**(stationary process)，相应的策略称为**平稳策略**(stationary policy)：

(1) 允许决策集合 $U_k(x)$与 $k$ 无关，于是可记作 $U(x)$，$x\in X$ 为状态变量；

(2) 状态转移 $T_k$ 与 $k$ 无关，于是可写作

$$x' = T(x,u), \tag{4.63}$$

$x$，$u$ 为当前阶段的状态和决策，$x'$为下一阶段的状态；

(3) 阶段指标 $v_k$ 与 $k$ 无关，可记作 $v(x,u)$.

不定期平稳过程的阶段变量 $k$ 视为任意的正整数.所谓不定期通常是给定状态的某一集合 $X_n$，称终止集合，当过程从 $k=1$ 的某状态 $x_1$ 出发按(4.63)转移时，一旦对于某个 $k=n$ 使得第 $n$ 段

的状态 $x_n \in X_n$，过程即告终结. 而优化问题是求最优决策序列，使指标函数 $V_{1n}$ 达到最小(大). 与最优策略相对应的 $n$ 称**最优历程**.

如果决策序列 $p_{1n}=\{u_1(x_1),\cdots,u_n(x_n)\}$ 中 $u_k$ 与 $k$ 无关，称为平稳的，可用一个函数 $u(x)$ 表示. 平稳过程的最优策略一定是平稳策略，记作 $p_{1n}^*=u^*(x)$.

不定期平稳过程的最优值函数 $f(x)$ 定义为

$$f(x)=\operatorname*{opt}_{n\geqslant 1,\,p_{1n}\in U(x)} V_{1n}(x,p_{1n}). \tag{4.64}$$

与定期决策过程的定理 4.2.1～4.2.3 相对应，不定期平稳过程也有最优策略 $p_{1n}^*$ 的一组充要条件. 下面只给出最优值函数 $f(x)$ 的基本方程.

**定理 4.4.6** 由(4.64)定义的 $f(x)$ 满足函数方程

$$f(x)=\operatorname*{opt}_{u\in U(x)}\left[v(x,u)+f(T(x,u))\right],x\in X.$$

过程终止条件为

$$f(x)=\varphi(x),\ x\in X_n\quad(\varphi\text{是已知函数}).$$

当最优解(4.64)存在且定理 4.4.6 函数方程的解 $f(x)$ 唯一时，它就是最优值函数，相应的 $u^*(x)$ 就是最优策略. 可以用函数迭代法或策略迭代法解这个函数方程.

函数迭代法的步骤是：

(1) 选初始函数 $f_0(x)$(一般取 $f_0(x)=0$).

(2) 用迭代公式

$$f_k(x)=\operatorname*{opt}_{u\in U(x)}\left[v(x,u)+f_{k-1}(T(x,u))\right],x\in X,$$

及 $f_k(x)=\varphi(x),x\in X_n$ 计算 $f_k(x),k=1,2,\cdots$.

(3) 当

$$f_{k+1}(x)=f_k(x),x\in X,$$

或

$$\left|\frac{f_{k+1}(x)-f_k(x)}{f_k(x)}\right|<\epsilon,\ x\in X\quad(\epsilon\text{是事先给定的数})$$

时迭代停止.最优值函数 $f(x)=f_k(x)$,最优策略 $u^*(x)=u_k(x)$.

策略迭代法的步骤是:

(1) 选初始策略 $u_1(x)$,令 $k=1$.

(2) 用 $u_k(x)$求解 $f_k(x)$,

$$f_k(x)=v(x,u_k(x))+f_k(T(x,u_k(x))),x\in X.$$

$$f_k(x)=\varphi(x),x\in X_n.$$

(3) 用 $f_k(x)$求改进策略 $u_{k+1}(x)$,

$$u_{k+1}(x)=(u|\operatorname*{opt}_{u\in U(x)}[v(x,u)+f_k(T(x,u))]).$$

(4) 当

$$u_{k+1}(x)=u_k(x),x\in X,$$

或

$$\left|\frac{f_{k+1}(x)-f_k(x)}{f_k(x)}\right|<\epsilon,x\in X$$

时迭代停止,最优值函数 $f(x)=f_k(x)$,最优策略 $u^*(x)=u_k(x)$;否则以 $k+1$ 代替 $k$ 重复(2),(3).

函数迭代法和策略迭代法中,序列$\{f_k(x)\}$和$\{u_k(x)\}$的收敛性在相当广泛的条件下是可以保证的,一般说来它与 $U(x)$,$T(x,u)$,$v(x,u)$,$X_n$ 等的具体形式有关.

### 4.4.3 无限期平稳过程

一个 $n$ 段决策过程要在 $n\to\infty$时有意义,即可以定义无限期过程,必须满足条件:

(1) 对于初始状态 $x_1$ 当允许决策集合 $U_1(x_1)\neq\varnothing$时,存在决策序列$\{u_1(x_1),u_2(x_2),\cdots\}$,使对于任意 $k$,允许决策集合 $U_k(x_k)\neq\varnothing$,且状态转移方程 $x_{k+1}=T_k(x_k,u_k)$有定义.

(2) 指标函数 $V_{1n}$对于任意的 $n$ 有定义,且$\lim\limits_{n\to\infty}V_{1n}$存在,记作 $V$.

优化问题可定义为求最优策略 $p^*$使

$$V(x,p^*) = \underset{p\in P}{\text{opt}}\, V(x,p),$$

$x$ 是初始状态，$P$ 是允许策略集合.

从应用角度只有无限期平稳过程才是有意义的. 若指标函数是阶段指标之和，则

$$V(x,p(x)) = v(x,u) + V(T(x,u),p(T(x,u))),$$

其中 $p(x)$是平稳策略，即 $p(x)=u(x)$.

引入无限期平稳过程的最优值函数

$$f(x) = \underset{p\in P}{\text{opt}}\, V(x,p(x)), \tag{4.65}$$

$f(x)$的基本方程由下述定理给出.

**定理 4.4.7** (4.65)定义的 $f(x)$满足函数方程

$$f(x) = \underset{u\in U(x)}{\text{opt}} [v(x,u) + f(T(x,u))], x \in X.$$

当最优解(4.65)存在且该函数方程有唯一解 $f(x)$时，它就是最优值函数，相应的 $u^*(x)$为最优策略.

无限期平稳过程的基本方程与不定期平稳过程在形式上是一样的，所以可以用函数迭代法或策略迭代法按照类似于例 4.4.2 中步骤求解.

无限期平稳过程可以用段数 $n$ 充分大的定期过程逼近，函数迭代法中 $f_n(x)$可选择为 $n$ 段的定期过程的最优值函数. 反之，段数 $n$ 很大的定期过程也可以视作无限期平稳过程，用求解函数方程代替计算递推方程.

## 4.5 随机性多阶段决策过程

在确定性决策过程中对于给定的初始状态和决策序列，按照状态转移方程得到的状态序列构成一条确定的轨线，在这条轨线上定义的指标函数是确定的数值. 随机性决策过程与确定性决策过程的区别在于：第一，过程演变时随机因素的存在使得状态转

移是以转移概率形式确定的，即使给定初始状态和初始决策，下一阶段的状态也不能确定，不能构成确定的轨线，最优策略应从每个阶段所有状态的决策函数集合中寻求. 第二，确定性过程中的指标函数已没有意义，随机性过程的指标函数应取它对于转移概率分布的期望值.

### 4.5.1 基本方程

随机性决策过程中最基本也是应用最广的是所谓**独立干扰过程**(indepedently distrubed process)，即状态转移方程为

$$x_{k+1} = T_k(x_k, u_k, w_k), \tag{4.66}$$

其中 $x_k \in X_k, u_k(x_k) \in U_k(x_k), k=1,2,\cdots,n$. 定义同确定性过程，$w_k \in W_k$ 为第 $k$ 段的随机干扰，由概率分布 $P_k(w_k | x_k, u_k)$ 表示其统计规律，$W_k$ 是随机干扰集合. 特别地，假定随机变量 $w_k$ 对于不同的 $k$ 相互独立.

对于独立干扰过程，由概率分布 $P_k(w_k | x_k, u_k)$ 和方程(4.66)可以确定状态转移概率，记作 $P_k(x_{k+1} | x_k, u_k)$，表示由 $x_k$ 出发在决策 $u_k$ 下转移到 $x_{k+1}$ 的条件概率.

原来的阶段指标 $v_k$ 与 $w_k$ 有关，记作 $v_k(x_k, u_k, w_k)$，仍设指标函数为阶段指标之和的形式，则随机性过程的指标函数定义为

$$V_{1n}(x_1, p_{1n}) \triangleq E_{(w_1,\cdots,w_n)}\left[\sum_{k=1}^{n} v_k(x_k, u_k, w_k)\right], \tag{4.67}$$

式中 $E$ 表示数学期望，$p_{1n}=\{u_1(x_1),\cdots,u_n(x_n)\}$，这里 $u_k(x_k)$ 是第 $k$ 段所有可能状态 $x_k$ 的决策函数.

类似地定义后部子过程的指标函数 $V_{kn}(x_k, p_{kn})$. 最优值函数为

$$f_k(x_k) = \underset{p_{kn} \in P_{kn}(x_k)}{\text{opt}} V_{kn}(x_k, p_{kn}). \tag{4.68}$$

与确定性过程(定理 4.2.3)对应，随机性过程有以下定理.

**定理 4.5.1** $\{f_k(x_k)\}$ 是最优值函数序列和 $\{u_k^*(x_k)\}$ 是最优

决策序列的充要条件是它们满足

$$f_k(x_k) = \underset{u_k \in U_k(x)}{\text{opt}} [E_{w_k}(v_k(x_k, u_k, w_k) + f_{k+1}(x_{k+1}))],$$

$$k = n, n-1, \cdots, 2, 1,$$

$$f_{n+1}(x_{n+1}) = \varphi(x_{n+1}) \quad (\varphi \text{为已知函数}),$$

其中 $x_{k+1}$ 由转移方程(4.66)给出.

由于状态转移的无后效性独立干扰过程是一类 **Markov 决策过程**(Markov decision process). 当随机变量 $w_k$ 不具有相互独立性时,可以用扩充状态变量的方法化为独立干扰过程.

### 4.5.2 几个典型问题

**例 4.5.2** 随机存储问题

重新考察例 4.3.1. 第 $k$ 周需求量 $w_k$ 为随机变量,概率分布 $P_k(w_k)$ 已知,且 $w_k(k=1,2,\cdots,n)$ 相互独立,其他条件不变. 建立随机性动态规划模型,求最优策略使 $n$ 周的总期望费用最小,并在 $P_k(w_k)$ 与 $k$ 无关(即同分布)的情况下讨论最优解的形式.

**解** 状态 $x_k$ 为周初存储量,决策 $u_k$ 为订货量,在需求 $w_k$ 下状态转移方程为

$$x_{k+1} = x_k + u_k - w_k. \tag{4.69}$$

$v_k$ 可表示为

$$\begin{aligned} v_k(x_k, u_k, w_k) = &(a + bu_k)\delta(u_k) + c_1 \max(x_k + u_k - w_k, 0) \\ &+ c_2 \max(w_k - x_k - u_k, 0), \end{aligned} \tag{4.70}$$

其中

$$\delta(u_k) = \begin{cases} 1, & u_k \neq 0, \\ 0, & u_k = 0. \end{cases}$$

按照定理 4.5.1 最优值函数 $f_k(x_k)$ 满足

$$f_k(x_k) = \min_{u_k \in U_k(x_k)} [E_{w_k} v_k(x_k, u_k, w_k) + E_{w_k} f_{k+1}(x_{k+1})],$$

$$U_k(x_k) = \{u_k | 0 \leqslant u_k \leqslant m_1 - x_k\}, k = n, \cdots, 2, 1, \tag{4.71}$$

$$f_{n+1}(x_{n+1}) = \varphi(x_{n+1}). \tag{4.72}$$

$m_1$ 是最大允许存储量，$\varphi$ 表示剩余库存的价值.

(4.71)，(4.72)即这个问题的动态规划模型. 当 $w_k$ 的概率分布 $P_k(w_k)$ 给定后，可以递推地求出最优值函数序列$\{f_k(x_k)\}$和最优决策序列$\{u_k^*(x_k)\}$.

下面设 $P_k(w_k)$同分布，且为简单起见设 $\varphi(x_{n+1})=0$. 由固定订货费 $a$ 等于零和不等于零分两种情况：

(1) $a=0$，由(4.70)～(4.72)可得

$$f_k(x_k) = \min_{u_k \in U_k(x_k)} [by_k + L(y_k) + E_w f_{k+1}(y_k - w_k)] - bx_k,$$

$$L(y) = c_1 E_w \max(y - w, 0) + c_2 E_w \max(w - y, 0),$$

$$y_k = x_k + u_k,$$

$$f_{n+1}(x_{n+1}) = 0. \tag{4.73}$$

可以证明最优策略为

$$u_k^*(x_k) = \begin{cases} S_k - x_k, & x_k < S_k, \\ 0, & x_k \geqslant S_k, \end{cases} \tag{4.74}$$

其中 $S_k$ 是函数

$$G_k(y) = by + L(y) + E_w f_{k+1}(y - w) \tag{4.75}$$

的极小值点.

(4.74)表明，最优策略由序列$\{S_k\}$给出，当第 $k$ 周初的存储量 $x_k$ 小于 $S_k$ 时，则订货，并使存储量达到 $S_k$；否则就不订货，这里假定 $S_k \leqslant m_1$.

(2) $a \neq 0$. 可以证明在(4.75)定义的 $G_k(y)$为 $K$-凸函数($K=a$)的条件下最优策略为

$$u_k^*(x_k) = \begin{cases} S_k - x_k, & x_k < s_k, \\ 0, & x_k \geqslant s_k \end{cases} \tag{4.76}$$

其中 $S_k$ 定义同上，而 $s_k$ 是使

$$G_k(y) = a + G_k(S_k) \tag{4.77}$$

成立的 $y$ 的最小值.

函数 $G(x)$ 称为 $K$-凸的($K>0$),如果对于任意的正数 $\alpha,\beta$ 有

$$K+G(x+\alpha)-G(x)-\frac{\alpha}{\beta}[G(x)-G(x-\beta)]\geqslant 0. \tag{4.78}$$

(4.77)表明,最优策略由两个序列 $\{s_k\}$ 和 $\{S_k\}$ 给出,当 $x_k<s_k$ 时则订货,并使存储量达到 $S_k$;否则不订货.这种策略称为 $(s,S)$ 型策略.

**例 4.5.3** 随机损坏下的设备更新问题

重新考察例 4.3.2.考虑一个设备损坏、必须更新的随机因素,确定性决策过程就变成随机性的.设一台年初年龄为 $t$ 的设备在年末损坏的概率为 $p(t)$,如果损坏,更新的净费用为 $q_1(t)$(新设备的购价扣除坏设备的回收).其它条件不变.建立动态规划模型,求最优策略使 $n$ 年的总期望收入最大.

**解** 参照例 4.3.2 中的基本方程(4.22),注意到阶段指标 $v_k$ 与随机因素无关,即 $E_{w_k}v_k=v_k$,而 $E_{w_k}J_{k+1}(x_{k+1})$ 可以根据概率 $p(t)$ 计算,最优值函数 $f_k(t)$ 的基本方程为

$$f_k(t)=\max\begin{bmatrix}r(0)-d(0)-q(t)+p(0)[-q_1(1)+f_{k+1}(0)]\\ \quad +[1-p(0)]f_{k+1}(1),\ u_k=R\\ r(t)-d(t)+p(t)[-q_1(t+1)+f_{k+1}(0)]\\ \quad +[1-p(t)]f_{k+1}(t+1),\ u_k=K\end{bmatrix},$$

$$k=n-1,\cdots,1. \tag{4.79}$$

为了写出终端条件,设第 $n$ 年末年龄为 $t$ 的好设备的回收为 $\varphi(t)$,坏设备的回收为 $\varphi_1(t)$,则

$$f_n(t)=\max\begin{bmatrix}r(0)-d(0)-q(t)+p(0)\varphi_1(1)+[1-p(0)]\varphi(1)\\ r(t)-d(t)+p(t)\varphi_1(t+1)+[1-p(t)]\varphi(t+1)\end{bmatrix}. \tag{4.80}$$

按照(4.80),(4.79)递推至 $f_1(s)$ 可以求出最优策略.

**例 4.5.4** 水电站水库调度问题

水电站水库调度是指根据观测资料和一定的规律决定通过发电机的水流流量,使发电量尽可能大.发电量与流量和水位的乘积成正比,水位取决于水库蓄水量,蓄水量又取决于入库径流,径流与雨量有关,可以看作以年为周期的随机过程.若将一年划分为若干阶段(如每旬或每月为一阶段),各阶段径流的概率分布可以由统计资料得到,那么水库调度通常是根据每阶段初的水位(可以换算成库容量)和径流的概率分布,决定本阶段的流量,使电站在尽可能高的水位上运行,达到年发电量的期望值最大的目的.试建立动态规划模型寻求最优调度方案.

**解** 假定入库径流具有最简单的性质:各阶段径流相互独立,第 $k$ 段径流 $w_k$ 的概率分布为 $P_k(w_k)$, $k=1,2,\cdots,n$(一年分成 $n$ 个阶段).实际上,只要适当地划分时段,大体上可以满足这种独立性.

取第 $k$ 段水库蓄水量为状态 $x_k$,泄放的发电流量为决策 $u_k$,则状态转移方程为

$$x_{k+1} = x_k + w_k - u_k, \tag{4.81}$$

第 $k$ 段的发电量与该段平均水位和流量 $u_k$ 的乘积成正比,平均水位取决于 $x_k$ 和 $x_{k+1}$, $x_{k+1}$ 由(4.81)给出,所以第 $k$ 段发电量可以记作 $v(x_k,u_k,w_k)$.

最优值函数 $f_k(x_k)$ 定义为从第 $k$ 段状态 $x_k$ 出发到年末的最大期望发电量,满足方程

$$\begin{aligned} f_k(x_k) &= \max_{u_k \in U(x_k)} E_{w_k}[v(x_k,u_k,w_k) + f_{k+1}(x_{k+1})] \\ &= \max_{u_k \in U(x_k)} \sum_{w_k} P_k(w_k)[v(x_k,u_k,w_k) + f_{k+1}(x_{k+1})], \\ & k = n,\cdots,2,1. \end{aligned} \tag{4.82}$$

$U(x_k)$由具体情况确定.

水电站是多年运行，最优调度要使多年平均的年期望发电量最大. 因为径流是以年为周期的随机过程，故对于不同的年而言，决策过程是平稳的，可以用函数迭代法求解，按下式计算：

$$f_k^{(i)}(x_k) = \max_{u_k \in U(x_k)} \sum_{w_k} P_k(w_k)\left[v(x_k, u_k, w_k) + f_{k+1}^{(i)}(x_{k+1})\right],$$

$$k = n, \cdots, 2, 1, \qquad i = 1, 2, \cdots.$$

$$f_{n+1}^{(1)}(x_{n+1}) = 0. \tag{4.83}$$

如记

$$g^{(i)}(x_1) = f_1^{(i)}(x_1) - f_1^{(i-1)}(x_1),$$

$$g_1^{(i)} = \min_{x_1} g^{(i)}(x_1),$$

$$g_2^{(i)} = \max_{x_1} g^{(i)}(x_1),$$

可以证明随着 $i$ 的增加，$g_1^{(i)}$和 $g_2^{(i)}$将分别单调增和单调减地收敛于最大的多年平均年期望发电量.

**例 4.5.5** 独立干扰线性系统的二次指标函数问题

在定理 4.3.10 确定性线性系统二次指标函数问题中考虑独立干扰因素，状态转移方程为

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{w}_k, \tag{4.84}$$

$\boldsymbol{w}_k$ 为相互独立的随机变量，给定概率分布且具有零均值和有限方差. 其它条件不变，求下列二次指标函数的极小值问题.

$$\min_{u_1, \cdots, u_n} E\sum_{k=1}^{n}(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q}_k\boldsymbol{x}_k + 2\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{u}_k + \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{R}_k\boldsymbol{u}_k) + \boldsymbol{x}_{n+1}^{\mathrm{T}}\boldsymbol{Q}_{n+1}\boldsymbol{x}_{n+1}. \tag{4.85}$$

最优值函数 $f_k(\boldsymbol{x})$满足

$$f_k(\boldsymbol{x}_k) = \min_{u_k} E_{w_k}\left[v_k(\boldsymbol{x}_k, \boldsymbol{u}_k) + f_{k+1}(\boldsymbol{x}_{k+1})\right], k = n, \cdots, 2, 1.$$

$$f_{n+1}(\boldsymbol{x}_{n+1}) = \boldsymbol{x}_{n+1}^{\mathrm{T}}\boldsymbol{Q}_{n+1}\boldsymbol{x}_{n+1}. \tag{4.86}$$

其中

$$v_k(\boldsymbol{x}_k,\boldsymbol{u}_k) = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{Q}_k\boldsymbol{x}_k + 2\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{S}_k\boldsymbol{u}_k + \boldsymbol{u}_k^{\mathrm{T}}\boldsymbol{R}_k\boldsymbol{u}_k.$$

可以证明,最优控制函数为

$$\begin{aligned}
&\boldsymbol{u}_k^*(\boldsymbol{x}_k) = -\boldsymbol{L}_k\boldsymbol{x}_k,\\
&\boldsymbol{L}_k = (\boldsymbol{R}_k + \boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)^{-1}(\boldsymbol{S}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)^{\mathrm{T}},\\
&\boldsymbol{P}_k = \boldsymbol{Q}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{A}_k - (\boldsymbol{S}_k + \boldsymbol{A}_k^{\mathrm{T}}\boldsymbol{P}_{k+1}\boldsymbol{B}_k)\boldsymbol{L}_k,\\
&\boldsymbol{P}_{n+1} = \boldsymbol{0}, \qquad\qquad k = 1,2,\cdots,n.
\end{aligned} \tag{4.87}$$

最优值函数为

$$f_k(\boldsymbol{x}_k) = \boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{P}_k\boldsymbol{x}_k + \sum_{j=k}^{n} E(\boldsymbol{w}_j^{\mathrm{T}}\boldsymbol{P}_{j+1}\boldsymbol{w}_j). \tag{4.88}$$

比较(4.87)与(4.54)可知其形式全然相同.由此得到重要结论:随机性过程的结果与确定性过程相同,只要在随机性过程中以 $E\boldsymbol{w}$ 代替 $\boldsymbol{w}$,就可以化为确定性过程.这个结论称**确定性等价原理**.

## 4.6 确定性连续决策过程

时间连续的决策过程优化问题通常称为最优控制.问题的典型提法是

$$\begin{aligned}
\min_{\boldsymbol{u}(t)}\quad & \int_0^T g(\boldsymbol{x}(t),\boldsymbol{u}(t)^*,t)\mathrm{d}t + h(\boldsymbol{x}(\tau),\tau);\\
\text{s.t.}\quad & \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{u}(t),t),\\
& \boldsymbol{x}(0) = \boldsymbol{x}_0,\boldsymbol{x}(t) \in X,\boldsymbol{u}(t) \in U.
\end{aligned} \tag{4.89}$$

其中 $\boldsymbol{x}(t)$是 $n$ 维状态函数,$\boldsymbol{u}(t)$是 $r$ 维策略函数,$\boldsymbol{f}$ 是相应维数的向量函数,$g$ 和 $h$ 是标量函数.

用动态规划方法求解(4.89),定义最优值函数 $J(\boldsymbol{x}(t),t)$为由时刻 $t$ 的状态 $\boldsymbol{x}(t)$出发至终点 $T$ 的目标函数的最优值,即

$$J(\boldsymbol{x}(t),t) = \min_{\boldsymbol{u}(\tau)}\int_t^T g(\boldsymbol{x}(\tau),\boldsymbol{u}(\tau),\tau)\mathrm{d}\tau + h(\boldsymbol{x}(T),T) \tag{4.90}$$

将连续过程离散化，利用动态规划基本方程，并对最优值函数作 Taylor 展开，再令时间间隔趋于零，可以得到 $J(\boldsymbol{x}(t),t)$满足的方程，有以下定理.

**定理 4.6.1** 设(4.89)的 $g,\boldsymbol{f}$ 连续可微，最优值函数 $J$ 二次可微，则 $J(\boldsymbol{x}(t),t)$满足 Hamilton-Jacobi-Bellman 方程(简称 HJB 方程)

$$-\frac{\partial J}{\partial t}=\min_{\boldsymbol{u}(t)}\left[g(\boldsymbol{x}(t),\boldsymbol{u}(t),t)+\left(\frac{\partial J}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{u}(t),t)\right]. \tag{4.91}$$

边界条件是

$$J(\boldsymbol{x}(T),T)=h(\boldsymbol{x}(T),T), \tag{4.92}$$

方程的解 $\boldsymbol{u}(t)$为最优控制函数.

若记

$$\begin{aligned}&H(\boldsymbol{x}(t),\boldsymbol{u}(t),\boldsymbol{\lambda}(t),t)\\&=g(\boldsymbol{x}(t),\boldsymbol{u}(t),t)+\boldsymbol{\lambda}^{\mathrm{T}}(t)\boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{u}(t),t),\end{aligned} \tag{4.93}$$

$$\boldsymbol{\lambda}(t)=\frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}(t),t). \tag{4.94}$$

HJB 方程可记作

$$-\frac{\partial J}{\partial t}=\min_{\boldsymbol{u}(t)}H(\boldsymbol{x}(t),\boldsymbol{u}(t),\boldsymbol{\lambda}(t),t), \tag{4.95}$$

$H$ 称 Hamilton 函数.

线性系统二次指标函数的最优控制问题(4.89)中的 $g,\boldsymbol{f},h$ 分别为

$$\left.\begin{aligned}g(\boldsymbol{x}(t),\boldsymbol{u}(t),t)&=\boldsymbol{x}(t)\boldsymbol{Q}(t)\boldsymbol{x}(t)+2\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{S}(t)\boldsymbol{u}(t)\\&\quad+\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{R}(t)\boldsymbol{u}(t),\\\boldsymbol{f}(\boldsymbol{x}(t),\boldsymbol{u}(t),t)&=\boldsymbol{A}(t)\boldsymbol{x}(t)+\boldsymbol{B}(t)\boldsymbol{u}(t),\\h(\boldsymbol{x}(T),T)&=\boldsymbol{x}^{\mathrm{T}}(T)L\boldsymbol{x}(T).\end{aligned}\right\} \tag{4.96}$$

其中 $\boldsymbol{R}(t)$为正定阵，$\boldsymbol{Q}(t),L$ 为半正定阵.

**定理 4.6.2** 线性系统二次指标函数最优控制问题(4.89),(4.96)的最优解为

$$u^*(x,t) = -R^{-1}(t)[S(t) + P(t)B(t)]^{\mathrm{T}}x(t), \tag{4.97}$$

$$\dot{P}(t) = -Q(t) - A^{\mathrm{T}}(t)P(t) - P(t)A(t) + [S(t) + P(t)B(t)]^{\mathrm{T}}R^{-1}(t)[S(t) + P(t)B(t)], \tag{4.98}$$

$$P(T) = L.$$

最优值函数为

$$J(x(t),t) = x^{\mathrm{T}}(t)P(t)x(t).$$

(4.98)是 Riccati 微分方程. 对于定常系统($A$,$B$ 为常数阵),并且指标函数中 $Q$,$S$,$R$ 也为常阵时(终端条件应去掉). 上述 Riccati 微分方程化为 Riccati 代数方程,其解 $P$ 也是常数阵. (4.97)表明,最优控制函数是状态的线性负反馈.

## 4.7 计算方法的改进

用动态规划模型可以求出全局最优解,是以在每个阶段对每个状态和每个决策用枚举法进行计算和比较作为代价的. 于是当状态变量的维数变大时存在着维数灾这一严重的缺点,目前尚没有克服这个困难的一般方法. 下面给出几种改进算法可供选用.

### 4.7.1 最优值函数近似法

用一维状态变量情形说明这种算法的基本思想.

对于每个阶段 $k$,若状态 $x$ 有 $m$ 个取值,则在基本方程递推过程中要计算并储存最优值函数 $f_k(x)$的 $m$ 个数值,$m$ 一般较大. 如果函数 $f_k$ 比较光滑,可以用一个低阶多项式近似,那么只要根据 $f_k(x)$的若干个数值(个数比 $m$ 小得多),确定出多项式的系数并储存下来,就可以为下一阶段($k-1$)计算最优值函数 $f_{k-1}(x)$提

供所要求的 $f_k(x)$ 的数值.

可以采用以下几种近似方法.

(1) 正交多项式近似

1) 一维情形

[−1,1]区间上的 Legendre 多项式(下称 L 多项式) $\{\varphi_i(x)\}$ 满足递推关系

$$\varphi_{i+1}(x)=\frac{2i+1}{i+1}x\varphi_i(x)-\frac{i}{i+1}\varphi_{i-1}(x),i=1,2,\cdots,$$

$$\varphi_0(x)=1,\varphi_1(x)=x.$$

及正交性质

$$\int_{-1}^{1}\varphi_i(x)\varphi_j(x)\mathrm{d}x=\begin{cases}\dfrac{2}{2i+1}, & i=j,\\ 0, & i\neq j.\end{cases}$$

任何连续平方可积函数 $f(x)$ 可以表为

$$f(x)=\sum_{i=0}^{\infty}a_i\varphi_i(x),$$

$$a_i=\frac{2i+1}{2}\int_{-1}^{1}f(x)\varphi_i(x)\mathrm{d}x.$$

有限和 $\sum_{i=0}^{r}a_i\varphi_i(x)$ 是 $f(x)$ 在[−1,1]上 $r$ 次多项式的最小平方近似,而 $a_i$ 可以利用 L 多项式的性质以求和代替积分近似地计算

$$a_i=\frac{2i+1}{2}\sum_{j=1}^{l}\lambda_j f(\mu_j)\varphi_i(\mu_j),$$

$$\lambda_j=\frac{2(1-\mu_j^2)}{(l+1)^2\varphi_{l+1}^2(\mu_j)},$$

其中 $\mu_j$ 是 $\varphi_l(x)=0$ 的第 $j$ 个根,$l$ 事先选定.

在最优值函数的递推计算中,先用变量代换将状态 $x$ 的变化范围转化到[−1,1]上. 设 $f_{k+1}(x)$ 用 L 多项式的 $r$ 次近似表示为

$$f_{k+1}(x) = \sum_{i=0}^{r} a_i^{(k+1)} \varphi_i(x), \tag{4.99}$$

其中 $a_i^{(k+1)}$ 已经确定，则 $f_k(x)$ 的 L 多项式 $r$ 次近似

$$f_k(x) = \sum_{i=0}^{r} a_i^{(k)} \varphi_i(x) \tag{4.100}$$

中的系数 $a_i^{(k)}$ 可以由

$$a_i^{(k)} = \frac{2i+1}{2} \sum_{j=1}^{l} \lambda_j f_k(\mu_j) \varphi_i(\mu_j),$$

$$f_k(\mu_j) = \underset{u_k}{\text{opt}} \left[ v_k(\mu_j, u_k) + \sum_{i=0}^{r} a_i^{(k+1)} \varphi_i(T_k(\mu_j, u_k)) \right] \tag{4.101}$$

递推地算出. 在选定 $r, l$ 以后，$\mu_j, \lambda_j$ 和 $\varphi_i(\mu_j)$ 都可以事先算好作为已知数据输入.

2) $n$ 维情形

$n$ 元函数 $f(x_1, \cdots, x_n)$ 的 L 多项式的 $r$ 次近似为

$$f(x_1, \cdots, x_n) = \sum_{i=0}^{r} a_i \prod_{s=1}^{n} \varphi_{p(i,s)}(x_s),$$

其中 $\prod_{s=1}^{n} \varphi_{p(i,s)}(x_s)$ 是 $n$ 个变元的所有 L 多项式的乘积，只要其积的次数不超过 $r$，即对于每个 $i$，$\sum_{s=1}^{n} p(i,s) \leqslant r$. 与一维情形类似，$a_i$ 可以近似地计算

$$a_i = \alpha_i \sum_{j_1=1}^{l} \cdots \sum_{j_n=1}^{l} \lambda_{j_1} \cdots \lambda_{j_n} f(\mu_{j_1}, \cdots, \mu_{j_n}) \prod_{s=1}^{n} \varphi_{p(i,s)}(\mu_{j_s}),$$

$\mu_{j_s}$ 是 $\varphi_l(x_s)=0(1 \leqslant s \leqslant n)$ 的第 $j_s$ 个根，$\lambda_{j_s}$ 是由 $\mu_{j_s}$ 和 $\varphi_l$ 确定的固定数值，$\alpha_i$ 是由 $i$ 确定的规格化常数.

利用这种近似完全可以将最优值函数中系数 $a_i^{(k)}$ 的递推公式 (4.99)～(4.101) 推广到 $n$ 维情形.

(2) 多项式回归

如果选定用 $r$ 次多项式近似最优值函数 $f_k(x)$，取状态 $x$ 的 $p$ 个值计算最优值函数 $f_k(x_j)$，$j=1,2,\cdots,p$，$p\gg r$. 然后用最小二乘法使 $\sum_{j=1}^{p}[f_k(x_j)-(a_0+a_1x_j+\cdots+a_rx_j^r)]^2$ 达到最小，确定出系数 $a_0,a_1,\cdots,a_r$，则多项式 $a_0+a_1x+\cdots+a_rx^r$ 可以作为最优值函数 $f_k(x)$ 的近似表达式供递推计算之用，而只需储存 $r+1$ 个系数.

(3) 分段近似

根据经验或粗略计算判断，最优值函数不光滑、甚至不连续时，可以按照具体情况将 $x$ 的变化范围分为若干个区间，在每个区间上用较低阶的多项式近似. 具体方法可以参考(1)，(2)，或者用**样条函数**(spline function)拟合.

(2)，(3)都能推广到多维的情形.

### 4.7.2 Lagrange 乘子法

这种方法用于将高维问题降低维数. 下面以例 4.3.5 给出的两种资源分配问题为例，说明用 Lagrange 乘子将二维问题化为一维问题求解的方法.

这个问题的静态模型(例 4.3.5)在引入乘子 $\lambda$ 后，可以写作

$$\begin{aligned}\max_{u_k,v_k}\quad & \left[\sum_{k=1}^{n}g_k(u_k,v_k)-\lambda\sum_{k=1}^{n}v_k\right];\\ \text{s.t.}\quad & \sum_{k=1}^{n}u_k=m_1.\end{aligned}\tag{4.102}$$

构造动态规划模型时将 $\lambda$ 视为参数，定义阶段指标

$$\tilde{g}_k(u_k,\lambda)=\max_{v_k}[g_k(u_k,v_k)-\lambda v_k].\tag{4.103}$$

于是状态 $x_k$ 和决策 $u_k$ 均降为一维变量，状态转移方程为

$$x_{k+1}=x_k-u_k.\tag{4.104}$$

最优值函数 $f_k(x_k,\lambda)$ 满足

$$f_k(x_k,\lambda) = \max_{v_k}[\tilde{g}_k(u_k,\lambda) + f_{k+1}(x_{k+1})], k = n,\cdots,2,1.$$

$$f_{n+1}(0,\lambda) = 0. \tag{4.105}$$

具体计算步骤是

(1) 任给 $\lambda \geqslant 0$.

(2) 解一维动态规划(4.103)～(4.104),计算至 $f_1(m_1,\lambda)$,设得到 $r$ 个最优策略,记作 $\{u_1^{(i)}(\lambda),u_2^{(i)}(\lambda),\cdots,u_n^{(i)}(\lambda)\}$, $i=1,2,\cdots,r$. 与 $u_k^{(i)}(\lambda)$ 相应的由(4.103)确定的 $v_k$ 的最优解记作 $\{v_1^{(i)}(\lambda),v_2^{(i)}(\lambda),\cdots,v_n^{(i)}(\lambda)\}$, $i=1,2,\cdots,r$.

(3) 若存在 $i$ 使

$$\sum_{k=1}^{n} v_k^{(i)}(\lambda) = m_2, \tag{4.106}$$

则 $\{u_k^{(i)}\}$ 和 $\{v_k^{(i)}\}$ 为原问题的最优解. 否则计算

$$F(\lambda) = \min_i \sum_{k=1}^{n} v_k^{(i)}(\lambda),$$

$$G(\lambda) = \max_i \sum_{k=1}^{n} v_k^{(i)}(\lambda).$$

(4) 若 $F(\lambda)>m_2$ 则增加 $\lambda$,转(2);若 $G(\lambda)<m_2$ 则减少 $\lambda$,转(2).

(5) 若 $F(\lambda)<m_2<G(\lambda)$,则将原问题的 $u_k$ 与 $v_k$ 对换($m_1$ 与 $m_2$ 也对换),重新回到(1). 若仍然有 $F(\lambda)<m_1<G(\lambda)$,则 Lagrange 乘子法失效.

上述步骤的正确性由以下定理保证.

**定理 4.7.1** 若 $\{u_k(\lambda),v_k(\lambda)\}$ 是(4.102)的最优解,则它也是原问题(4.30)、(4.31)的最优解,只是(4.31)中的 $\sum_{k=1}^{n} v_k = m_2$ 代之以 $\sum_{k=1}^{n} v_k = \sum_{k=1}^{n} v_k(\lambda)$.

**定理 4.7.2** 若 $\lambda_1>\lambda_2$，则 $G(\lambda_1)\leqslant F(\lambda_2)$.

由定理 4.7.1 一旦(4.106)成立(或近似成立)则得原问题最优解. 由定理 4.7.2 当 $\lambda$ 增加时 $\sum_{k=1}^{n} v_k(\lambda)$ 下降(非增)，于是 $F(\lambda)$ 减小. 同样，当 $\lambda$ 减小时 $\sum_{k=1}^{n} v_k(\lambda)$ 上升(非降)，于是 $G(\lambda)$ 增加. 这给出了(4)中修正 $\lambda$ 求解的依据. 同时也表明在(5) $F(\lambda)<m_2<G(\lambda)$ 的情况下，不论增加或减小 $\lambda$ 都无法得到满足(4.106)的最优解.

这种方法可以推广到高于二维的情况，如对于 $s$ 维的分配问题

$$
\begin{aligned}
&\max_{u_{kj}} \quad \sum_{k=1}^{n} g_k(u_{k1},u_{k2},\cdots,u_{ks});\\
&\text{s.t.} \quad \sum_{k=1}^{n} u_{kj}=m_j,\ j=1,2,\cdots,s.
\end{aligned}
\tag{4.107}
$$

引入 $p(<s)$ 个 Lagrange 乘子 $\lambda_j(j=1,2,\cdots,p)$，将(4.107)写作

$$
\begin{aligned}
&\max_{u_{kj}} \quad \left[\sum_{k=1}^{n} g_k(u_{k1},\cdots,u_{ks})-\sum_{j=1}^{p}\lambda_j\sum_{k=1}^{n}u_{kj}\right];\\
&\text{s.t.} \quad \sum_{k=1}^{n} u_{kj}=m_j,\ j=p+1,\cdots,s.
\end{aligned}
\tag{4.108}
$$

对于一组给定的 $\lambda_1,\cdots,\lambda_p$，用 $s-p$ 维动态规划模型解(4.108)，其步骤与(1)～(5)类似.

### 4.7.3 几种实用算法

(1) 逐次逼近法

求多元函数极值的一种数值方法是，轮流让一个变元变动而其它变元都固定在已得到的最优值上，求解一元函数的极值问题，从而逐次地逼近多元函数的极值点. 这就是**逐次逼近法**(succes-

sive approximation method). 它可以用来求解状态为多维变量的动态规划.

设状态 $x_k$ 是 $s$ 维向量,记作 $x_k=(x_{k1},x_{k2},\cdots,x_{ks})$,在第 $j$ 次逼近时已得到最优轨线为 $x_k^{(j)}=(x_{k1}^{(j)},\cdots,x_{ks}^{(j)})$,$k=1,2,\cdots,n$. 在第 $j+1$ 次逼近中令

$$x_{kl}=\begin{cases}x_{kl}^{(j)},l=p+1,\cdots,s,\\ x_{kl}^{(j+1)},l=1,\cdots,p-1.\end{cases}$$

求解 $s$ 个一维状态变量 $x_{kp}$ 的动态规划,$p$ 依次由 1 变到 $s$,从而得到 $x_k^{(j+1)}$. 同时得到相应的最优策略 $u_k^{(j+1)}$,$k=1,2,\cdots,n$.

这种方法可以大大减小存储空间,当然计算时间会相应地增加.

(2) 松弛法

如果根据实际情况对最优轨线有一个初步估计 $x^\circ=\{x_1^\circ,x_2^\circ,\cdots,x_n^\circ\}$,则可将原来的允许状态集合 $X=\{X_1,X_2,\cdots,X_n\}$ 缩小为以 $x^\circ$ 为中心的某个邻域 $X^\circ=\{X_1^\circ,X_2^\circ,\cdots,X_n^\circ\}$,于是允许决策集合也相应地缩小为 $U^\circ=\{U^\circ{}_1,U^\circ{}_2,\cdots,U^\circ{}_n\}$,$U^\circ{}_k=\{u_k|u_k\in U_k(x_k),T_k(x_k,u_k)\in X_{k+1}^\circ\}$. $X^\circ$ 和 $U^\circ$ 将显著地小于原来的 $X$ 和 $U$.

对于 $X^\circ$ 和 $U^\circ$ 求出最优策略 $p^\circ=\{u_1^\circ,u_2^\circ,\cdots,u_n^\circ\}$ 和最优轨线 $x^1=\{x_1^1,x_2^1,\cdots,x_n^1\}$,若 $x^1$ 是 $X^\circ$ 的内点,即所有的 $x_k^1$ 都是 $X^\circ{}_k$ 的内点($k=1,2,\cdots,n$),则表明附加约束 $X^\circ$ 不起作用,$p^\circ$ 为原问题的最优策略;否则,某些 $x_k^1$ 出现在 $X_k^\circ$ 的边界上,则以 $x^1$ 为中心构成新的邻域 $X^1$ 和相应的 $U^1$,重复上述过程直到某一步 $s$,得到的最优轨线 $x^s=\{x_1^s,x_2^s,\cdots,x_n^s\}$ 是 $X^{s-1}$ 的内点,这时最优策略 $p^{s-1}=\{u_1^{s-1},\cdots,u_n^{s-1}\}$ 为原问题的最优策略. 这种方法称为**松弛法**(relaxation method).

(3) 局部加密法

对于状态先取较大的格点间隔(疏网格),求出较粗糙的最优解,在最优轨线附近缩小格点间隔(密网格),再求最优解. 如此继

续下去直到求得满意的最优解为止.

这种**局部加密法**(local refined method)适用于最优值函数比较平滑的情况.

(4) 状态增量法

这是一种解连续决策过程(4.89)的离散化方法.

普通求解(4.89)的离散化方法是将时间区间$[0,T]$固定地分为 $N$ 段,于是每段长度为 $\Delta t=T/N$. 状态转移方程可写作

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{f}(\boldsymbol{x}_k,\boldsymbol{u}_k,k)\Delta t,k=1,2,\cdots,N.$$

最优值函数 $F_k(\boldsymbol{x}_k)$满足

$$F_k(\boldsymbol{x}_k)=\min_{\boldsymbol{u}_k}[g(\boldsymbol{x}_k,\boldsymbol{u}_k,k)\Delta t+F_{k+1}(\boldsymbol{x}_{k+1})],$$

这里状态 $\boldsymbol{x}_k$ 是 $n$ 维向量,略去下标 $k$ 记作 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$, $\boldsymbol{f}$ 是 $n$ 维向量函数,记作 $\boldsymbol{f}=(f_1,f_2,\cdots,f_n)$.

**状态增量法**(state increment method)的基本思想是每段时间长度 $\Delta t$ 不固定,在状态转移平缓时令 $\Delta t$ 大些,而在状态转移剧烈时令 $\Delta t$ 小些,使得每阶段状态转移都控制在较小的范围内,从而大大减少每阶段最优值函数的计算和存储量. 具体作法如下.

首先将状态 $\boldsymbol{x}$ 的各个分量 $x_l,l=1,2,\cdots,n$ 离散化,记 $x_l$ 的离散单位为 $\Delta x_l$. 设允许决策集合为 $U(\boldsymbol{x})=\{u^{(1)},u^{(2)},\cdots,u^{(s)}\}$,令

$$\Delta t^{(j)}=\min_{1\leqslant l\leqslant n}\left|\frac{\Delta x_l}{f_l(\boldsymbol{x},\boldsymbol{u}^{(j)},t)}\right|,\quad j=1,2,\cdots,s. \tag{4.109}$$

状态转移方程为

$$\boldsymbol{x}^{(j)}=\boldsymbol{x}+\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u}^{(j)},t)\Delta t^{(j)}, \tag{4.110}$$

则在时刻 $t$ 由状态 $\boldsymbol{x}$ 出发的最优值函数 $F(\boldsymbol{x},t)$满足的基本方程为

$$F(\boldsymbol{x},t)=\min_{1\leqslant j\leqslant s}[g(\boldsymbol{x},\boldsymbol{u}^{(j)},t)\Delta t^{(j)}+F(\boldsymbol{x}^{(j)},t+\Delta t^{(j)})]. \tag{4.111}$$

由(4.109)、(4.110)可知,在由 $\boldsymbol{x}$ 到 $\boldsymbol{x}^{(j)}$的转移中,通常只有一个分量变化一个离散单位 $\Delta x$,其它各分量的变化均不足一个单

位，在计算(4.111)时，$F(x^{(j)},t+\Delta t^{(j)})$需要在格点上进行扦值.

(5) 微分动态规划

**微分动态规划**(differential dynamic programming)是求解连续决策过程(4.89)的最优性方程(即 HJB 方程)的一种近似方法.对最优值函数在最优轨线附近作 Taylor 展开，略去高阶项，得到相应系数的常微分方程.用迭代法求解并保证每次迭代使目标函数值下降.

将问题(4.89)的最优值函数重新记作

$$\varphi(\boldsymbol{x},t)=\min_{u}\left[\int_t^T g(\boldsymbol{x},\boldsymbol{u},t)\mathrm{d}t+h(\boldsymbol{x}(T),T)\right],$$
$$\varphi(\boldsymbol{x},T)=h(\boldsymbol{x}(T),T), \tag{4.112}$$

$\varphi(\boldsymbol{x},t)$满足的最优性方程为

$$-\frac{\partial\varphi}{\partial t}=\min_u H(\boldsymbol{x},\boldsymbol{u},t,\varphi_x),$$
$$H(\boldsymbol{x},\boldsymbol{u},t,\varphi_x)=g(\boldsymbol{x},\boldsymbol{u},t)+\varphi_x^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x},\boldsymbol{u},t). \tag{4.113}$$

最优策略 $\boldsymbol{u}^*$ 和最优轨线 $\boldsymbol{x}^*$ 记作

$$\boldsymbol{u}^*=\bar{\boldsymbol{u}}+\delta\boldsymbol{u},$$
$$\boldsymbol{x}^*=\bar{\boldsymbol{x}}+\delta\boldsymbol{x}.$$

且 $\boldsymbol{x}=\bar{\boldsymbol{x}}$ 的目标值函数为

$$\psi(\bar{\boldsymbol{x}},t)=\int_t^T g(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},t)\mathrm{d}t+h(\bar{\boldsymbol{x}}(T),T).$$

当策略由 $\bar{\boldsymbol{u}}$ 变为 $\boldsymbol{u}^*$ 时目标值的改变为

$$a(\bar{\boldsymbol{x}},t)=\varphi(\bar{\boldsymbol{x}},t)-\psi(\bar{\boldsymbol{x}},t)\leqslant 0. \tag{4.114}$$

$\varphi(\bar{\boldsymbol{x}}+\delta\boldsymbol{x},t)$在 $\bar{\boldsymbol{x}}$ 展开后略去高阶项，并利用(4.114)得

$$\varphi(\bar{\boldsymbol{x}}+\delta\boldsymbol{x},t)=\psi(\bar{\boldsymbol{x}},t)+a(\bar{\boldsymbol{x}},t)$$
$$+\varphi_x^{\mathrm{T}}\delta\boldsymbol{x}+\frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}\varphi_{xx}\delta\boldsymbol{x}. \tag{4.115}$$

将(4.115)代入最优性方程(4.113)得

$$
\begin{aligned}
&-\frac{\partial\psi}{\partial t}-\frac{\partial a}{\partial t}-\frac{\partial\varphi_x^{\mathrm{T}}}{\partial t}\delta x-\frac{1}{2}\delta x^{\mathrm{T}}\frac{\partial\varphi_{xx}}{\partial t}\delta x \\
&=\min_{\delta u}[H(\bar{x}+\delta x,\bar{u}+\delta u,t,\varphi_x) \\
&\quad+(\varphi_{xx}\delta x)^{\mathrm{T}}f(\bar{x}+\delta x,\bar{u}+\delta x,t)].
\end{aligned}
\tag{4.116}
$$

设 $u$ 是 $H$ 的极小点,满足$\frac{\partial H}{\partial u}(\bar{x},u,t,\varphi_x)=0$. $u$ 对应的状态为 $x$,设 $x$ 在 $\bar{x}$ 附近,即 $\delta x=x-\bar{x}$ 很小. 最优策略 $u^*=u+\delta u^*$,而 $\delta u^*$ 对应的 $\delta x^*\doteq\delta x$.

将(4.116)的右端在$(\bar{x},u)$处展开,略去高阶项后得到关于 $\delta u^*$ 的二次式. 设 $\delta u^*=\beta\delta x^*$,可解出 $\beta$,得

$$
\delta u^*=-H_{uu}^{-1}(H_{ux}+f_u^{\mathrm{T}}\varphi_{xx})\delta x^*,
\tag{4.117}
$$

且(4.116)化为(注意到 $\delta x=\delta x^*$)

$$
\begin{aligned}
&-\frac{\partial\psi}{\partial t}-\frac{\partial a}{\partial t}-\frac{\partial\varphi_x^{\mathrm{T}}}{\partial t}\delta x^*-\frac{1}{2}\delta x^{*\mathrm{T}}\frac{\partial\varphi_{xx}}{\partial t}\delta x^* \\
&=H+(H_x+\varphi_{xx}f)\delta x^*+\frac{1}{2}\delta x^{*\mathrm{T}}[(H_{xx}+f_x^{\mathrm{T}}\varphi_{xx}+\varphi_{xx}f_x) \\
&\quad-(H_{ux}+f_u^{\mathrm{T}}\varphi_{xx})^{\mathrm{T}}H_{uu}^{-1}(H_{ux}+f_u^{\mathrm{T}}\varphi_{xx})]\delta x^*.
\end{aligned}
$$

比较两端 $\delta x^*$ 同次幂系数得

$$
\begin{cases}
-\dfrac{\partial\psi}{\partial t}-\dfrac{\partial a}{\partial t}=H, \\
-\dfrac{\partial\varphi_x}{\partial t}=H_x+\varphi_{xx}f \\
-\dfrac{\partial\varphi_{xx}}{\partial t}=(H_{xx}+f_x^{\mathrm{T}}\varphi_{xx}+\varphi_{xx}f_x) \\
\qquad\qquad-(H_{ux}+f_u^{\mathrm{T}}\varphi_{xx})^{\mathrm{T}}H_{uu}^{-1}(H_{ux}+f_u^{\mathrm{T}}\varphi_{xx})
\end{cases}
\tag{4.118}
$$

注意到

$$
\frac{\mathrm{d}\varphi}{\mathrm{d}t}=\frac{\partial\varphi}{\partial t}+\varphi_x^{\mathrm{T}}f(\bar{x},\bar{u},t),
$$

$$
\frac{\mathrm{d}a}{\mathrm{d}t}=\frac{\mathrm{d}\varphi}{\mathrm{d}t}-\frac{\mathrm{d}\psi}{\mathrm{d}t}.
$$

(4.118)可化为常微分方程组

$$\begin{cases} \dfrac{\mathrm{d}a}{\mathrm{d}t} = H(\bar{x},\bar{u},t,\varphi_x) - H, \\ \dfrac{\mathrm{d}\varphi_x}{\mathrm{d}t} = \varphi_{xx}[f(\bar{x},\bar{u},t) - f] - H_x, \\ \dfrac{\mathrm{d}\varphi_{xx}}{\mathrm{d}t} = (H_{ux} + f_u^{\mathrm{T}}\varphi_{xx})^{\mathrm{T}}H_{uu}^{-1}(H_{ux} + f_u^{\mathrm{T}}\varphi_{xx}) \\ \qquad\qquad - (H_{xx} + f_x^{\mathrm{T}}\varphi_{xx} + \varphi_{xx}f_x). \end{cases} \tag{4.119}$$

式中除指明的外均在$(\bar{x},u)$处计算.端点条件为

$$\begin{cases} a(\bar{x},T) = 0, \\ \varphi_x(T) = h_x(\bar{x}(T),T), \\ \varphi_{xx}(T) = h_{xx}(\bar{x}(T),T). \end{cases} \tag{4.120}$$

对于给定的策略$\bar{u}$和由方程$\dot{x}=f(x,u,t)$得到的相应的状态$\bar{x}$,可以算出目标值函数$\psi(\bar{x},t)$.对$\bar{u}$作微小改变,如只在小区间$[t_1,T]$($t_1$靠近$T$)上将$\bar{u}$改为$u$($u$是$H(\bar{x},u,t,\varphi_x)$的极小值点),使相应的状态$\delta x$改变也很小.判断$\delta x$是否是小量的标准为检查

$$\frac{\Delta\psi(\bar{x},t)}{|a(\bar{x},t)|} \geqslant c \tag{4.121}$$

是否成立,$c$常取0.5,其中$\Delta\psi$是目标值函数的实际改进量,而$|a|$是策略变为最优时的预期改变量.$a$的计算显然需要求解(4.119).若(4.121)不成立,可改变$t_1$和$c$的值.如此进行的迭代每次都使目标函数值下降,具体步骤如下.

(1) 给定初始策略$\bar{u}$,解出相应的状态$\bar{x}$,算出目标值函数$\psi(\bar{x},t)$,给定小量$a_0$.

(2) 求$H$的极小值点$u$,解(4.119),(4.120)得到$a(\bar{x},t)$,$\varphi_x$,$\varphi_{xx}$.检查并记录$|a(\bar{x},t)|\geqslant a_0$的时刻$t'$.

(3) 若$t'=0$,即$|a(\bar{x},t)|<a_0$,$0\leqslant t\leqslant T$,则$u$即为所求,$u^*=u$;否则转(4).

(4) 令$u=u+\beta\delta x$,($t_0\leqslant t\leqslant T$,$t_0$可取0)其中$\beta=-H_{uu}^{-1}(H_{ux}$

$+f_u^T\varphi_{xx})$，$\delta x=x-\bar{x}$，而 $x$ 是 $\dot{x}=f(x,u+\beta\delta x,t)$的解. 计算这个策略 $u$ 下的目标值函数 $\psi'(\bar{x},t)$.

(5) 计算 $\Delta\psi(\bar{x},t)=\psi-\psi'$，检查$\dfrac{\Delta\psi(\bar{x},t)}{|a(\bar{x},t)|}\geqslant c(c=0.5)$是否成立，若成立，则以 $u$ 代替 $\bar{u}$，转(1)；否则转(6).

(6) 取 $t_{k+1}=t_k+\dfrac{1}{2}(T-t_k)$，$k=1,2,\cdots$，转(4)，重新计算 $\psi'$. 若对于这样的 $t_k$，$\dfrac{\Delta\psi(\bar{x},t)}{|a(\bar{x},t)|}\geqslant c$ 不能成立，减少 $c$，再进行计算. 若仍不行，则该方法失效.

按照上述步骤迭代，直到在 $0\leqslant t\leqslant T$ 满足 $|a(\bar{x},t)|<a_0$ 为止.

# 参考文献


1 王日爽，徐兵，魏权龄编. 应用动态规划. 国防工业出版社，1987

2 张润琦编. 动态规划. 北京理工大学出版社，1989

3 Bellman R E. Dynamic Programming. Princeton University Press, Princeton, New Jersey, 1957

4 Bellman R E and Dreyfus S E. Applied Dynamic Programming. Princeton University Press, Princeton, New Jersey, 1962

5 Denardo Eric V. Dynamic Programming, Models and Application. Prentice-Hall, INC., Englewood Cliffs, New Jersey, 1982

6 Dreyfus S E and Law A M. The Art and Theory of Dynamic Programming. Academic Press, INC., 1977

7 Jacobson D H and Mayne D Q. Differential Dynamic Programming. Elserier, New York, 1970

8 Larson R E and Casti J L. Principles of Dynamic Programming, Part Ⅱ. Marcel Dekker, INC. 1982

# 5　多目标规划

## 5.1　引言

**多目标规划**(multiobjective programming)是运筹学的一个分支，是近年来发展起来的新学科. 它研究在一定的约束条件下多个目标函数的极值问题. 而线性规划、非线性规划与整数规划等研究的都是单目标函数的极值问题. 然而很多实际问题是多目标的，而且有些目标互相冲突，因此要研究多目标规划. 如

**例 5.1.1**　某厂生产两种产品 $A$ 与 $B$，产品 $A$ 是畅销产品，有较高的声誉，供不应求，每月至少生产 1 万件. 但因原料限制，月产量不能超过 3 万件，每件产品 $A$ 的利润为 1 元. 生产一件 $B$ 所需工时与生产一件 $A$ 所需工时相同，但每件产品 $B$ 的利润为 3 元. 在工人不加班的正常情况下，每月两种产品产量总和不低于 5 万件，也不高于 7 万件. 产量超过 5 万件的部分叫做超产量. 工厂希望: (1) $A$ 的产量要多；(2) 为了使工人不加班，超产量要少；(3) 利润要多. 试建立该问题的数学模型.

**解**　设 $A$ 的月产量为 $x_1$ 万件，$B$ 的月产量为 $x_2$ 万件. 此问题的三个目标可以写成 $\max x_1$, $\min(x_1+x_2-5)$ 与 $\max(x_1+3x_2)$. 为了整齐，将第二个目标改写为 $\max(5-x_1-x_2)$，则此问题的数学模型为

$$
\begin{aligned}
\max\quad & [f_1(x_1,x_2), f_2(x_1,x_2), f_3(x_1,x_2)];\\
\text{s.t.}\quad & x_1+x_2\geqslant 5,\\
& x_1+x_2\leqslant 7,\\
& \qquad x_1\geqslant 1,
\end{aligned}
$$

$$x_1 \leqslant 3,$$
$$x_2 \geqslant 0.$$

其中

$$f_1(x_1,x_2) = x_1,$$
$$f_2(x_1,x_2) = 5 - x_1 - x_2,$$
$$f_3(x_1,x_2) = x_1 + 3x_2.$$

这是多目标规划问题.

多目标规划在水资源利用、环境保护、计划管理、财政预算、工程设计及农业规划等方面有广泛的应用.

在多目标规划中,一般不存在所有目标函数共同的极大点. 例如,在图 5.1 中,$x_1$ 是 $f_1(x)$的极大点,$x_2$ 是 $f_2(x)$的极大点,但 $f_1(x)$与 $f_2(x)$没有共同的极大点. 因此需要引进**非劣解**(noninferior solution)的概念(将在 5.2 节介绍). 一般说,一个多目标规划有无穷多个非劣解. 使决策者满意的非劣解叫**最终解**(final solution). 5.3,5.4,5.5 节将介绍求非劣解与最终解的方法. 5.6 节介绍目标规划.

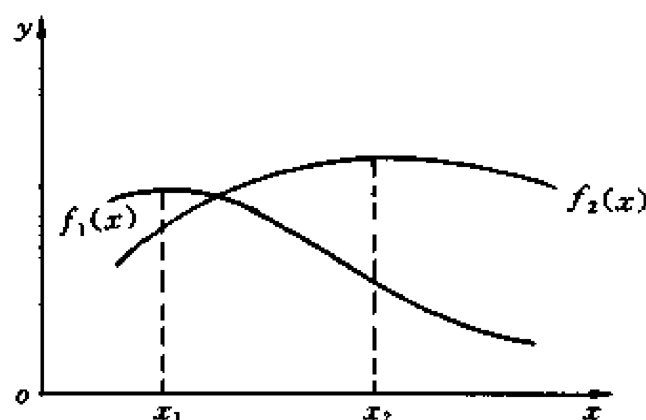



图 5.1

多目标规划的另一个特点是解法很多. 要根据实际情况,通过试算,选用合适的解法.

## 5.2 多目标规划的基本概念与 K-T 条件

### 5.2.1 多目标规划的基本概念

多目标规划的标准形式是

$$(\text{VOP}) \quad \max \quad [f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]; \tag{5.1}$$

$$\text{s.t.} \quad g_i(\boldsymbol{x}) \leqslant 0, i=1,\cdots,m. \tag{5.2}$$

其中 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 是 $n$ 维向量，$\boldsymbol{x}$ 所在的空间叫**决策空间**(decision space)，$f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})$ 称为目标函数，$p$ 维向量 $(f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x}))$ 所在的空间称为**目标空间**(objective space)，$g_1(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x})$ 称为约束函数. 多目标规划问题又称为**向量最优化问题**(vector optimization problem)，简记为(VOP).

例 5.1.1 是一个多目标规划问题，它的数学模型可改写为标准形式(5.1)，(5.2).

下面引进一些基本概念.

**定义 5.2.1** 集

$$X=\{\boldsymbol{x}\in E^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,\cdots,m\}$$

叫做多目标规划在**决策空间上的可行集**(feasible set in decision space).

**定义 5.2.2** 集

$$F=\{(f_1,\cdots,f_p)\in E^p \mid f_i=f_i(\boldsymbol{x}), i=1,\cdots,p, \boldsymbol{x}\in X\}$$

叫做多目标规划在**目标空间上的可行集**(feasible set in objective space).

应用记号 $X$，(5.1)与(5.2)所表示的多目标规划的标准形式可以改写为

$$(\text{VOP}) \quad \max \quad [f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]. \tag{5.3}$$

$$\text{s.t.} \quad x \in X \tag{5.4}$$

将向量函数 $\boldsymbol{f}(x)=[f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]$看成是映射 $\boldsymbol{f}:X\subset E^n\to E^p$,则目标空间上的可行集 $F$ 是决策空间上的可行集 $X$ 在映射 $\boldsymbol{f}$ 下的象(image),即 $F=\boldsymbol{f}(X)$. 如图 5.2 所示.

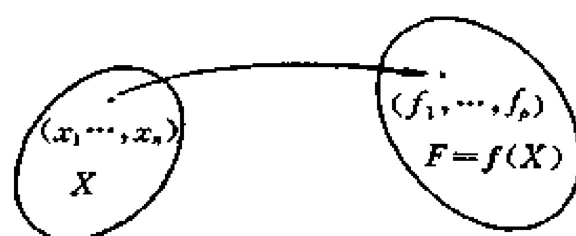



图 5.2

**定义 5.2.3** 若 $\boldsymbol{x}'\in X$,且不存在另一个可行点 $\boldsymbol{x}\in X$,使 $f_i(\boldsymbol{x})\geqslant f_i(\boldsymbol{x}')$,$i=1,\cdots,p$ 成立,且其中至少有一个严格不等式成立,则称 $\boldsymbol{x}'$是多目标规划(VOP)的一个**非劣解**(noninferior solution). 所有非劣解构成的集叫**非劣集**(noninferior set).

非劣解又称为**有效解**(efficient solution)、**非优超解**(nondominated solution)或 **Pareto 最优解**(Pareto optimal solution).

**定义 5.2.4** 若 $\boldsymbol{x}'\in X$,且不存在另一可行点 $\boldsymbol{x}\in X$,使 $f_i(\boldsymbol{x})>f_i(\boldsymbol{x}')$,$i=1,\cdots,p$ 成立,则称 $\boldsymbol{x}'$是多目标规划(VOP)的一个**弱非劣解**(weak noninferior solution). 所有弱非劣解构成的集叫**弱非劣集**(weak noninferior set).

**定义 5.2.5** 若 $\boldsymbol{x}'$是非劣解(弱非劣解),则称$[f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]$为**目标空间中的非劣解(弱非劣解)**.

在无混淆的情况下,目标空间中的非劣解(弱非劣解)也简称为非劣解(弱非劣解).

**定理 5.2.6** 若 $\boldsymbol{x}'$是非劣解,则一定是弱非劣解.

**例 5.2.7** 设两目标规划为

$$\max\ (f_1(\boldsymbol{x}),f_2(\boldsymbol{x}));$$

$$\text{s.t.}\ \boldsymbol{x} \in X = \{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \boldsymbol{x}^4\}.$$

这里 $X$ 是由 4 个点组成的可行集，$f_1(\boldsymbol{x})$与 $f_2(\boldsymbol{x})$的函数值由下表给出，求这个问题的非劣解与弱非劣解.

| $\boldsymbol{x}^i$ | $\boldsymbol{x}^1$ | $\boldsymbol{x}^2$ | $\boldsymbol{x}^3$ | $\boldsymbol{x}^4$ |
|---|---|---|---|---|
| $f_1(\boldsymbol{x}^i)$ | 9 | 10 | 12 | 9 |
| $f_2(\boldsymbol{x}^i)$ | 11 | 11 | 10 | 8 |
| 点$(f_1, f_2)$ | $A$ | $B$ | $C$ | $D$ |

**解** 因为 $f_1(\boldsymbol{x}^2) > f_1(\boldsymbol{x}^4)$，$f_2(\boldsymbol{x}^2) > f_2(\boldsymbol{x}^4)$，所以 $\boldsymbol{x}^4$ 不是弱非劣解，也不是非劣解. 按定义，$\boldsymbol{x}^2$ 与 $\boldsymbol{x}^3$ 都是非劣解. $\boldsymbol{x}^1$ 是弱非劣解. 因为非劣解一定是弱非劣解，$\boldsymbol{x}^2$ 与 $\boldsymbol{x}^3$ 也是弱非劣解. 非劣集是$\{\boldsymbol{x}^2, \boldsymbol{x}^3\}$，弱非劣集是$\{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3\}$. 图 5.3 给出此问题的目标空间的图形.

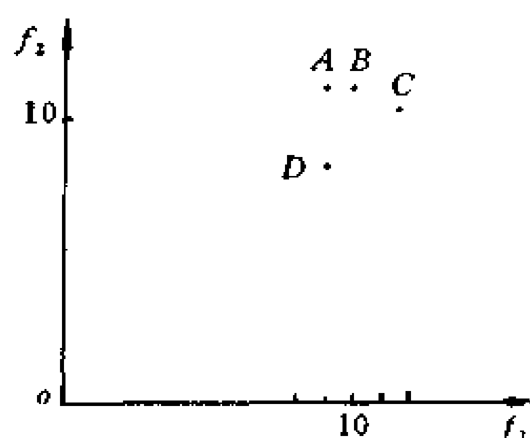



图 5.3

**非劣解的几何解释** 有两个目标函数时，目标空间中的非劣解有明显的几何解释. 在图 5.4 中，设 $F$ 是目标空间中的可行集. 可以看出，$F$ 的任一内部点 $P(f_1, f_2)$一定不是非劣解，因为在 $P$ 点东北侧(图中画阴影部分)的点 $P'$，它的坐标$(f_1', f_2')$必满足：

$f_1' \geqslant f_1, f_2' \geqslant f_2$，且其中至少有一个严格不等式成立. 如果 $F$ 中某点的东北侧是空集，则此点是非劣解. 因此 $F$ 的东北角边界(图中的 $BA$ 弧)上任一点是非劣解. $BA$ 弧是非劣集. 这里的非劣解与非劣集都是指目标空间中的非劣解与非劣集.

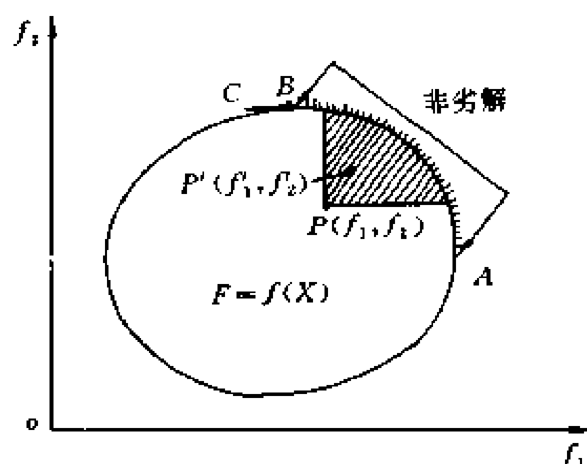



图 5.4

可以看出，$F$ 上部边界的水平线段 $CB$ 上的点是弱非劣解.

**定义 5.2.8** 若 $\boldsymbol{x}'$ 是多目标规划(VOP)的非劣解，且存在正数 $M$，使对满足 $f_i(\boldsymbol{x}) > f_i(\boldsymbol{x}')$ 的每一个 $i$ 与每一个 $\boldsymbol{x} \in X$，至少存在一个 $j \neq i$，使 $f_j(\boldsymbol{x}) < f_j(\boldsymbol{x}')$，与 $\dfrac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}')}{f_j(\boldsymbol{x}) - f_j(\boldsymbol{x}')} \geqslant -M$ 成立，则称 $\boldsymbol{x}'$ 是(VOP)的一个**真非劣解**(proper noninferior solution).

图 5.5 给出问题

$$\begin{aligned} \max \quad & [f_1(x), f_2(x)]; \\ \text{s.t.} \quad & x \geqslant 0 \end{aligned}$$

的目标函数 $f_1(x)$ 与 $f_2(x)$ 的图形，从图中可以看出，闭区间 $[x_1, x_2]$ 上任一点是非劣解. 可以验证开区间 $(x_1, x_2)$ 内任一点是真非劣解. 但两个端点 $x_1$ 与 $x_2$ 不是真非劣解，例如 $x_1$，我们设 $x = x_1 + h$，$h > 0$，当 $h \to 0^+$ 时，$\dfrac{f_2(x) - f_2(x_1)}{f_1(x) - f_1(x_1)} \to -\infty$，说明 $x_1$ 不是真非劣

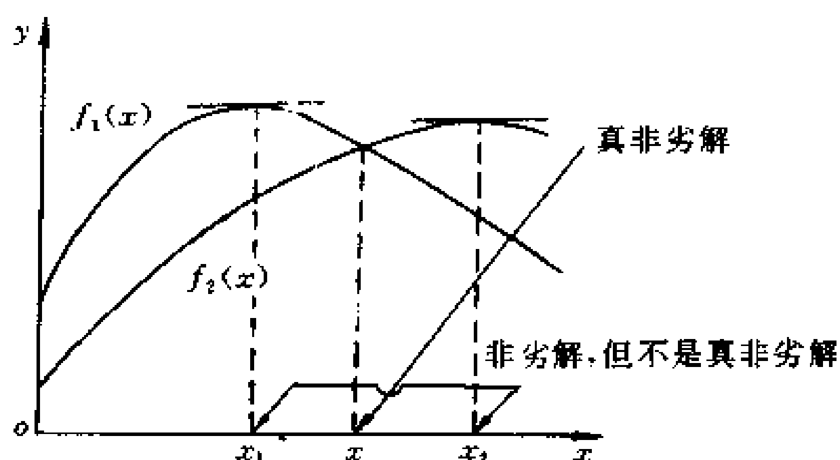



图 5.5

解.

**定义 5.2.9** 决策者对目标函数进行评价的函数 $U(f_1,\cdots,f_p)$ 叫做**效用函数**(utility function).

效用函数值的大小反映决策者对目标值的喜爱程度,若对两组目标值 $(f_1,\cdots,f_p)$ 与 $(f'_1,\cdots,f'_p)$ 有 $U(f_1,\cdots,f_p)>U(f'_1,\cdots,f'_p)$ 成立,表示决策者喜欢前者.

**定义 5.2.10** 使效用函数取最大值的非劣解称为**最佳协调解**(best compromise solution).

在许多实际问题中,不知道决策者效用函数的表达式,因此求不出最佳协调解.但通过分析者与决策者的对话,可取得决策者效用函数的信息,从而求出决策者满意的近似最佳协调解.

**定义 5.2.11** 使多目标规划决策过程满意结束的解叫**最终解**(final solution).

**定义 5.2.12** 效用函数 $U(f_1,\cdots,f_p)$ 的等值面 $U(f_1,\cdots,f_p)=c$ 称为**无差异曲面**(indifference surface),当 $p=2$ 时称为**无差异曲线**(indifference curve).

当 $p=2$ 时,求最佳协调解要解下述规划

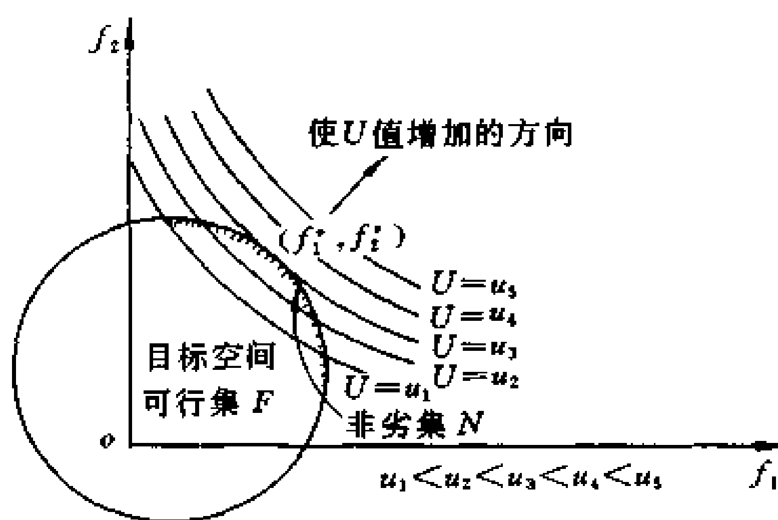



图 5.6

$$\max \quad U(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}));$$
$$\text{s.t.} \quad \boldsymbol{x} \in X = \{\boldsymbol{x} \in E^n \mid g_i(\boldsymbol{x}) \leqslant 0, i = 1, \cdots, m\}.$$

图 5.6 给出效用函数 $U(f_1, f_2)$ 的极大点代表最佳协调解的几何解释. 图中 $F=f(X)$ 是多目标规划在目标空间上的可行集, $N$ 是目标空间上的非劣集. 画出一系列无差异曲线 $U(f_1, f_2)=u_i$, 这里 $u_i$ 都是常数, $u_i$ 的值愈大表示决策者满意程度愈高. 对在同一条无差异曲线上不同的点, 决策者的满意程度相同. 当某条无差异曲线 $U(f_1, f_2)=u_i$ 与非劣集 $N$ 相切于(或只相交于)一点 $(f_1^*, f_2^*)$ 时, 相应的非劣解 $\boldsymbol{x}^*$ $(f_1^*=f_1(\boldsymbol{x}^*), f_2^*=f_2(\boldsymbol{x}^*))$ 就是最佳协调解.

**定义 5.2.13** 若 $A=(\hat{f}_1, \cdots, \hat{f}_p)$ 与 $B=(\hat{f}_1, \cdots, \hat{f}_l+\Delta\hat{f}_l, \cdots, \hat{f}_j+\Delta\hat{f}_j, \cdots, \hat{f}_p)$ 是一个曲面上的两个点, 若 $B$ 沿此曲面趋向于 $A$ 时, $-\dfrac{\Delta\hat{f}_l}{\Delta\hat{f}_j}$ 的极限存在, 则称 $t_{lj}=\lim\limits_{B\to A}\left(-\dfrac{\Delta\hat{f}_l}{\Delta\hat{f}_j}\right)$ 为在此曲面上目标 $f_l$ 对目对 $f_j$ 在点 $A$ 的**交换率**(trade-off rate).

在无差异曲面上的交换率叫边际替代率, 在经济学上很有用.

**定义 5.2.14** 在无差异曲面$U(f_1,\cdots,f_p)=U(\hat{f}_1,\cdots,\hat{f}_p)$上一点$A(\hat{f}_1,\cdots,\hat{f}_p)$的交换率$m_{lj}=\lim\limits_{B\to A}\left(-\dfrac{\Delta\hat{f}_l}{\Delta\hat{f}_j}\right)$称为在点$A$的**边际替代率**(marginal rate of substitution).

边际替代率的含意是,为了保持决策者同等的满意程度,要使$f_j$减少一个单位,$f_l$必须增加$m_{lj}$个单位.边际替代率又可以表示成

$$m_{lj}=\left.\frac{\dfrac{\partial U}{\partial f_j}}{\dfrac{\partial U}{\partial f_l}}\right|_{(\hat{f}_1,\cdots,\hat{f}_p)}\qquad l,j=1,\cdots,p\quad l\neq j.\tag{5.6}$$

若在$(\hat{f}_1,\cdots,\hat{f}_p)$的一个邻域内,无差异曲面可以表示成$f_l=f_l(f_1,\cdots,f_{l-1},f_{l+1},\cdots,f_p)$,则边际替代率可以写成

$$m_{lj}=-\left.\frac{\partial f_l}{\partial f_j}\right|_{(\hat{f}_1,\cdots,\hat{f}_{l-1},\hat{f}_{l+1},\cdots,\hat{f}_p)},\quad l,j=1,\cdots,p,j\neq l.\tag{5.7}$$

**边际替代率的几何解释**

在$p=2$时,无差异曲线$f_2=f_2(f_1)$在$(\hat{f}_1,\hat{f}_2)$点的边际替代率$m_{21}=-\left.\dfrac{df_2}{df_1}\right|_{f_1=\hat{f}_1}$是无差异曲线在此点切线斜率的负值(图5.7).

在图5.6中,$(f_1^*,f_2^*)$为最佳协调解的条件是无差异曲线与非劣集曲线在此点相切.该条件也可以叙述为:在此点非劣集的交换率等于边际替代率,即$t_{21}=m_{21}$.

**定义 5.2.15** 设$x^j$是

$$\begin{aligned}&\max\quad f_j(x),\\&\text{s.t.}\quad g_i(x)\leqslant 0,i=1,\cdots,m\end{aligned}$$

的最优解,相应的最优目标值是$M_j=f_j(x^j)$,$j=1,\cdots,p$,则称目

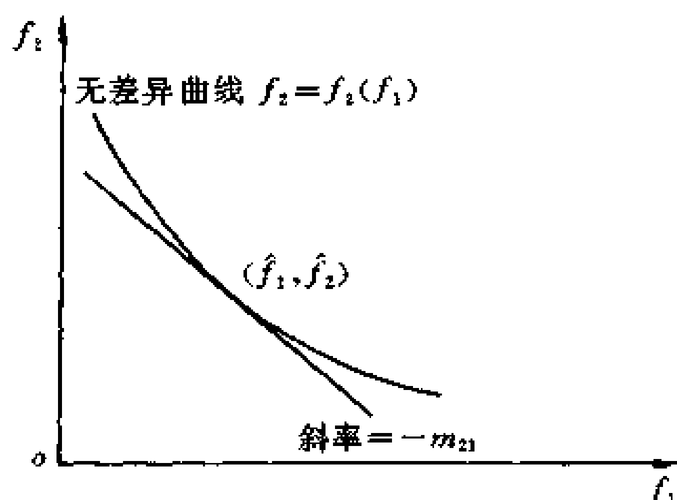



图 5.7

标空间中的点$(M_1,\cdots,M_p)$是多目标规划(5.1)、(5.2)的**理想点**(ideal point).

理想点是目标空间中每个目标都取极大值的点,在 $p=2$ 时,图 5.8 中画出理想点$(M_1,M_2)$的位置.

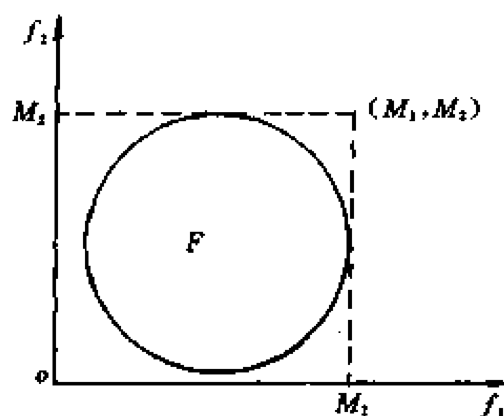



图 5.8

从图中可以看出,一般情况下,理想点$(M_1,M_2)\notin F$,是不可行的.也就是说,在现实情况下,不可能找到一个可行点 $x$ 使所有的目标函数都取极大值.

有时候，需要在某个 $l_\alpha$ 范数 $(1\leqslant\alpha<\infty)$ $d_\alpha=\left[\sum_{i=1}^{p}|f_i-f_i^*|^\alpha\right]^{1/\alpha}$ 下求距理想点最近的点，即求

$$\begin{aligned}&\max\quad \left[\sum_{i=1}^{p}|M_i-f_i(\boldsymbol{x})|^\alpha\right]^{1/\alpha},\\&\text{s.t.}\quad \boldsymbol{x}\in X\end{aligned}$$

的最优点 $\hat{\boldsymbol{x}}$ 与相应的函数值 $(f_1(\hat{\boldsymbol{x}}),\cdots,f_p(\hat{\boldsymbol{x}}))$.

可以证明，$\hat{\boldsymbol{x}}$ 是多目标规划(VOP)的非劣解，求这种非劣解的方法叫理想点法.

### 5.2.2 多目标规划的 K-T 条件

在非线性规划中有最优解的 K-T 条件(即 Kuhn-Tucker 条件)，类似地，在多目标规划中有关于非劣解的 K-T 条件.

设多目标规划为

$$\text{(VOP)}\qquad \begin{aligned}&\max\quad [f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})],\\&\text{s.t.}\quad g_i(\boldsymbol{x})\leqslant 0, i=1,\cdots,m.\end{aligned}$$

**定义 5.2.16** 若 $\boldsymbol{x}^*$ 是(VOP)的可行点，且梯度向量集 $\{\nabla g_i(\boldsymbol{x}^*)|i\in I\}$ 线性无关，其中 $I=\{i|g_i(\boldsymbol{x}^*)=0,i=1,\cdots,m\}$，则称 $\boldsymbol{x}^*$ 是可行集 $X=\{\boldsymbol{x}\in E^n|g_i(\boldsymbol{x})\leqslant 0,i=1,\cdots,m\}$ 的**正则点**(regular point).

**定理 5.2.17** 真非劣解的 Kuhn-Tucker 必要条件 设 $f_j(j=1,\cdots,p)$ 与 $g_i(i=1,\cdots,m)$ 是连续可微函数，$\boldsymbol{x}^*$ 是 $X$ 的正则点，则 $\boldsymbol{x}^*$ 是(VOP)的真非劣解的必要条件是：存在 $\lambda_j$ 与 $\mu_i$ 使

$$\left.\begin{aligned}&(1)\ \sum_{j=1}^{p}\lambda_j\nabla f_j(\boldsymbol{x}^*)-\sum_{i=1}^{m}\mu_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0};\\&(2)\ \mu_i g_i(\boldsymbol{x}^*)=0, i=1,\cdots,m;\\&(3)\ \lambda_j>0, j=1,\cdots,p, \mu_i\geqslant 0, i=1,\cdots,m.\end{aligned}\right\}\qquad(5.8)$$

成立.

**定理 5.2.18** 真非劣解的 Kuhn-Tucker 充分条件 设 $-f_j(j=1,\cdots,p)$与 $g_i(i=1,\cdots,m)$是连续可微的凸函数,则定理 5.2.17 的条件也是真非劣解的充分条件.

**定理 5.2.19** 非劣解的 Kuhn-Tucker 必要条件 设 $f_j(j=1,\cdots,p)$与 $g_i(i=1,\cdots,m)$是连续可微函数,且至少对一个 $k$,$\boldsymbol{x}^*$ 是问题 $P_k(\boldsymbol{\varepsilon}_k)$(参看(5.17)、(5.18))的一个正则点,则 $\boldsymbol{x}^*$ 是(VOP)的非劣解的必要条件是:存在 $\lambda_j$ 与 $\mu_i$ 使

$$\left.\begin{array}{l}(1)\ \sum_{j=1}^{p}\lambda_j\nabla f_j(\boldsymbol{x}^*)-\sum_{i=1}^{m}\mu_i\nabla g_i(\boldsymbol{x}^*)=\boldsymbol{0};\\ (2)\ \mu_i g_i(\boldsymbol{x}^*)=0;\\ (3)\ \lambda_j\geqslant 0,j=1,\cdots,p,\text{且至少有一个 }\lambda_j>0,\\ \quad\ \mu_i\geqslant 0,i=1,\cdots,m.\end{array}\right\}\tag{5.9}$$

成立.

**定理 5.2.20** 非劣解的 Kuhn-Tucker 充分条件 设 $-f_j(j=1,\cdots,p)$与 $g_i(i=1,\cdots,m)$是连续可微的凸函数,且 $g_i(i=1,\cdots,m)$是严格凸的,则 $\boldsymbol{x}^*$ 是(VOP)的非劣解的充分条件是:存在 $\lambda_j$ 与 $\mu_i$ 使(5.9)成立.

**定理 5.2.21** 弱非劣解的 Kuhn-Tucker 必要条件 设 $f_j(j=1,\cdots,p)$与 $g_i(i=1,\cdots,m)$是连续可微函数,$\boldsymbol{x}^*$ 是 $X$ 的正则点,则 $\boldsymbol{x}^*$ 是(VOP)的弱非劣解的必要条件是:存在 $\lambda_j$ 与 $\mu_i$ 使(5.9)成立.

**定理 5.2.22** 弱非劣解的 Kuhn-Tucker 充分条件 设 $-f_j(j=1,\cdots,p)$与 $g_i(i=1,\cdots,m)$是连续可微的凸函数,则定理 5.2.21 的条件也是弱非劣解的充分条件.

## 5.3 寻求多目标规划非劣解的方法

一个多目标规划如果存在非劣解,往往存在无穷多个,形成非

劣集.求解实际问题时,很多非劣解是无法应用的,还要找出使决策者满意的最终解(定义 5.2.11).求最终解有两类方法.一类是求非劣解的生成法,即先求出大量的非劣解,构成非劣集的一个子集,然后按照决策者的意图找出最终解.另一类为交互法,不先求很多非劣解,而是通过分析者与决策者对话的方式,逐步求出最终解.本节介绍生成法.

### 5.3.1 加权法

**加权法**(weighting method)将求解多目标规划的非劣解问题化为求解若干单目标规划问题.设多目标规划为

$$(\text{VOP}) \qquad \max \quad [f_1(\boldsymbol{x}), \cdots, f_p(\boldsymbol{x})];$$
$$\text{s.t.} \quad g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, \cdots, m.$$

可行集为 $X = \{\boldsymbol{x} \in E^n \mid g_i(\boldsymbol{x}) \leqslant 0, i = 1, \cdots, m\}$.加权法先给定权因子 $W_j \geqslant 0, j = 1, \cdots, p$,构造相应的单目标规划

$$(\text{WP}) \qquad \max \quad \sum_{j=1}^{p} W_j f_j(\boldsymbol{x});$$
$$\text{s.t.} \quad \boldsymbol{x} \in X.$$

在一定的条件下,(WP)的最优解是(VOP)的非劣解,系统地改变权因子$(w_1, \cdots, w_p)$的值,求解一系列(WP)问题,得到(VOP)的大量非劣解,构成近似的非劣集.

**定理 5.3.1** 若(1)$w_j > 0,\ j = 1, \cdots, p$;

或(2)$w_j \geqslant 0,\ j = 1, \cdots, p$,且(WP)有唯一解,则(WP)的最优解是(VOP)的非劣解.

**定理 5.3.2** 若 $w_j \geqslant 0, j = 1, \cdots, p$,则(WP)的最优解是(VOP)的弱非劣解.

从定理 5.3.2 可以看出,当 $w_j \geqslant 0, j = 1, \cdots, p$ 时,(WP)的最优解 $\boldsymbol{x}^*$ 不一定是(VOP)的非劣解.为了进一步判断 $\boldsymbol{x}^*$ 是不是非劣解,可用下述定理.

**定理 5.3.3** 若对于 $w_j \geqslant 0, j=1,\cdots,p, x^*$ 是(WP)的最优解，且当

(1) $\bar{\delta}=0$ 时，$x^*$ 是(VOP)的非劣解；

(2) $\bar{\delta}>0$ 时，$x^*$ 不是(VOP)的非劣解.

其中 $\bar{\delta}$ 是下述辅助规划问题的最优函数值.

$$\bar{\delta} = \max \delta = \sum_{j=1}^{p} \delta_j; \tag{5.10}$$

$$\left.\begin{aligned} \text{s.t. } & f_j(x) - \delta_j = f_j(x^*) \quad j=1,\cdots,p, \\ & x \in X, \\ & \delta_j \geqslant 0, j = 1,\cdots,p. \end{aligned}\right\} \tag{5.11}$$

加权法的几何解释：

当 $p=2$ 时，多目标规划是

$$\begin{aligned} \max \quad & [f_1(x), f_2(x)]; \\ \text{s.t.} \quad & x \in X = \{x \in E^n \mid g_i(x) \leqslant 0, i = 1,\cdots,m\}. \end{aligned}$$

在图 5.9 中画出目标空间中的可行集

$$F = \{(f_1, f_2) \mid f_1 = f_1(x), \; f_2 = f_2(x), x \in X\}.$$

$F$ 的东北角边界是非劣集 $N$. 给定权因子 $w_1>0, w_2>0$，作直线族 $w_1 f_1 + w_2 f_2 = c$，其中 $c$ 是参数，选择这样的 $c$ 使族中每条直线都与 $F$ 相交，斜率都是 $-w_1/w_2$，其中有一条 $c$ 值最大并与 $N$ 相切的直线，此直线与 $N$ 相切的点 $(f_1^*, f_2^*)$ 是多目标规划(VOP)的一个非劣解. 变动 $w_1$ 与 $w_2$ 的值，可以得到不同的平行直线族，从而得到不同的非劣解.

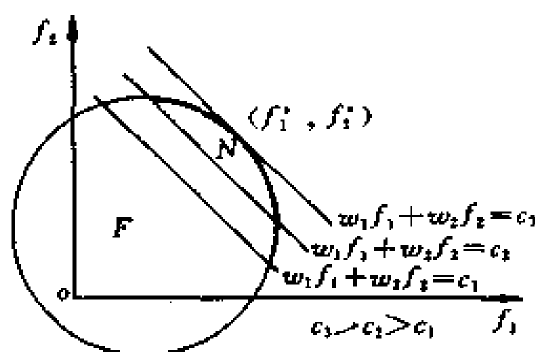



图 5.9

**例 5.3.4** 用加权法求下述问题的非劣解

$$\max \quad [f(x_1,x_2), f_2(x_1,x_2)] \tag{5.12}$$

$$\text{s.t.} \quad \left.\begin{aligned} -x_1+x_2 &\leqslant 4,\\ x_1+x_2 &\leqslant 9,\\ x_1 \quad &\leqslant 5,\\ x_2 &\leqslant 5,\\ -x_1 \quad &\leqslant 0,\\ -x_2 &\leqslant 0. \end{aligned}\right\} \tag{5.13}$$

其中

$$f_1(x_1,x_2) = 2x_1 - x_2,$$
$$f_2(x_1,x_2) = -x_1 + 3x_2$$

**解** 记可行集为

$$X = \{(x_1,x_2) \in E^2 \mid -x_1 + x_2 \leqslant 4, x_1 + x_2 \leqslant 9, 0 \leqslant x_1, x_2 \leqslant 5\},$$

目标函数的加权和为 $w_1f_1+w_2f_2$，相应的加权和单目标规划为

$$\max \quad w_1f_1(x_1,x_2) + w_2f_2(x_1,x_2); \tag{5.14}$$

$$\text{s.t.} \quad (x_1,x_2) \in X. \tag{5.15}$$

其中 $w_1,w_2>0$ 为加权因子. 此加权问题的最优解$(x_1,x_2)$应满足 Kuhn-Tucker 条件：对应于可行集的 6 个约束条件(见(5.13))存在 6 个 Lagrange 乘子 $\mu_i(i=1,\cdots,6)$，使

$$\frac{\partial}{\partial x_i}[w_1f_1(x_1,x_2) + w_2f_2(x_1,x_2)] - \sum_{j=1}^{6}\mu_j\frac{\partial}{\partial x_i}g_j(x_1,x_2) = 0,\ i = 1,2.$$

$$\mu_jg_j(x_1,x_2) = 0,\ j = 1,\cdots,6.$$

成立，即

$$\left.\begin{array}{r}2w_1 - w_2 + \mu_1 - \mu_2 - \mu_3 + \mu_5 = 0, \\ -w_1 + 3w_2 - \mu_1 - \mu_2 - \mu_4 + \mu_6 = 0, \\ (-x_1 + x_2 - 4)\mu_1 = 0, \\ (x_1 + x_2 - 9)\mu_2 = 0, \\ (x_1 - 5)\mu_3 = 0, \\ (x_2 - 5)\mu_4 = 0, \\ x_1\mu_5 = 0, \\ x_2\mu_6 = 0, \\ \mu_1, \cdots, \mu_6 \geqslant 0.\end{array}\right\} \quad (5.16)$$

成立. 因为(5.14)、(5.15)是一个线性规划,最优解在可行集的边界上. 又因有两个自变量,最优解有一个或两个起作用约束(参考定义 2.1.41). 先考虑只有 $x_2 \leqslant 5$ 起作用,此时等式 $x_2 = 5$ 成立,其他 5 个约束严格不等式成立,因此 $\mu_1 = \mu_2 = \mu_3 = \mu_5 = \mu_6 = 0$(由(5.16)推出). 解(5.16),得到 $w_1 = 1/3, w_2 = 2/3, \mu_4 = 5/3$. 用此组$(w_1, w_2)$值代入(5.14)、(5.15),并求解这个线性规划,得到无穷多个最优解,这些解分布在图 5.10 的 $CD$ 线段上. 这些最优解都是(5.12)、(5.13)的非劣解. 研究另外两个起作用约束的情况,得到

| 起作用约束 | Lagrange 乘子 | 解(5.16)得 | 非劣解 |
|---|---|---|---|
| $x_2 \leqslant 5$ | $\mu_1, \mu_2, \mu_3, \mu_5, \mu_6 = 0$ | $(w_1, w_2) = \left(\frac{1}{3}, \frac{2}{3}\right), \mu_4 = \frac{5}{3}$ | $CD$ 线段 |
| $x_1 + x_2 \leqslant 9$ | $\mu_1, \mu_3, \mu_4, \mu_5, \mu_6 = 0$ | $(w_1, w_2) = \left(\frac{4}{7}, \frac{3}{7}\right), \mu_2 = \frac{5}{7}$ | $DE$ 线段 |
| $x_1 \leqslant 5$ | $\mu_1, \mu_2, \mu_4, \mu_5, \mu_6 = 0$ | $(w_1, w_2) = \left(\frac{3}{4}, \frac{1}{4}\right), \mu_3 = \frac{5}{4}$ | $EF$ 线段 |

分析其他情况没有得到新的非劣解.

因此,折线$\overline{CDEF}$是非劣集,如图 5.10 与图 5.11 所示.

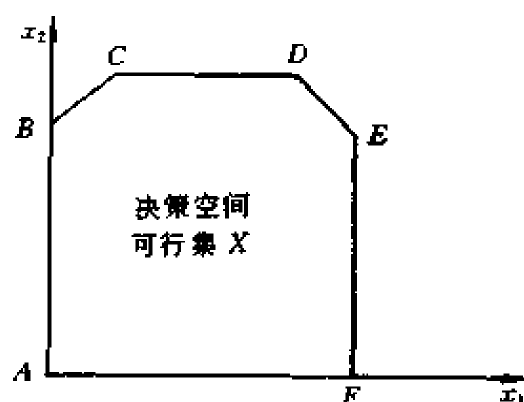



图 5.10

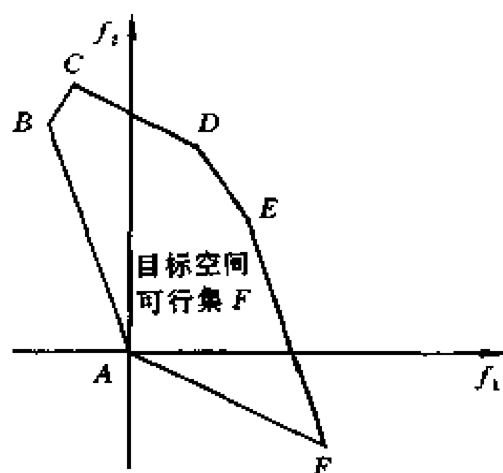



图 5.11

### 5.3.2 约束法

**约束法**(constraint method)也是一种用单目标规划求多目标规划非劣解的方法.设多目标规划为

(VOP) $\max \quad [f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]$;

$\text{s.t.} \quad g_i(\boldsymbol{x}) \leqslant 0, i=1,\cdots,m.$

可行集为 $X=\{\boldsymbol{x}\in E^n \mid g_i(\boldsymbol{x})\leqslant 0, i=1,\cdots,m\}$.选一个目标作为主目标,将其余目标变为约束,构造下述单目标规划问题

$$\mathrm{P}_k(\boldsymbol{\varepsilon}_k) \qquad \max f_k(\boldsymbol{x}); \tag{5.17}$$

$$\left.\begin{aligned} \text{s.t. } & f_j(\boldsymbol{x}) \geqslant \varepsilon_j, j=1,\cdots,p, j\neq k, \\ & \boldsymbol{x}\in X. \end{aligned}\right\} \tag{5.18}$$

这里 $\boldsymbol{\varepsilon}_k=(\varepsilon_1,\cdots,\varepsilon_{k-1},\varepsilon_{k+1},\cdots,\varepsilon_p)$是一个向量,其分量都是给定的常数,用 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$表示这个问题依赖于主目标函数的下标 $k$ 与约束参数 $\boldsymbol{\varepsilon}_k$.问题 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$称为原问题(VOP)的一个约束问题.关于 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$的最优解与(VOP)的非劣解之间的关系,有下述定理.

**定理 5.3.5** $\boldsymbol{x}^*$ 是多目标规划(VOP)的非劣解的充要条件是:$\boldsymbol{x}^*$ 同时是 $p$ 个约束问题 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k^*)(k=1,\cdots,p)$的最优解,这里 $\boldsymbol{\varepsilon}_k^*=$

$(\epsilon_1^*, \cdots, \epsilon_{k-1}^*, \epsilon_{k+1}^*, \cdots, \epsilon_p^*)$，其中 $\epsilon_j^* = f_j(\boldsymbol{x}^*)$，$j=1,\cdots,p$，$j \neq k$.

**定理 5.3.6** 对某一个 $k$，若 $\boldsymbol{x}^*$ 是 $P_k(\epsilon_k^*)$ 的唯一最优解，则 $\boldsymbol{x}^*$ 是多目标规划(VOP)的非劣解，这里 $\epsilon_k^* = (\epsilon_1^*, \cdots, \epsilon_{k-1}^*, \epsilon_{k+1}^*, \cdots, \epsilon_p^*)$，其中 $\epsilon_j^* = f_j(\boldsymbol{x}^*)$，$j=1,\cdots,p$，$j \neq k$.

根据这个定理，我们可以系统地改变 $\epsilon_j$ 的值，得到若干个约束问题 $P_k(\epsilon_k)$，分别求他们的最优解，可以得到多目标规划的许多非劣解. 若这些非劣解的个数足够多，可以构成近似的非劣集.

当 $p=2$ 时，约束问题

$$\max \quad f_2(\boldsymbol{x}); \tag{5.19}$$

$$\text{s.t.} \quad \left.\begin{array}{l} f_1(\boldsymbol{x}) \geqslant \epsilon_1, \\ \boldsymbol{x} \in X. \end{array}\right\} \tag{5.20}$$

的最优解有几何解释. 在图 5.12 中，$F = f(X)$ 是

$$\max[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})]$$
$$\text{s.t.}\ \boldsymbol{x} \in X$$

在目标空间中的可行集，$N$ 为非劣集，画阴影的区域是 $F$ 中满足 $f_1 \geqslant \epsilon_1$ 的部分，也是(5.19)、(5.20)在目标空间中的可行集. 从图上可以看出，点 $A$ 的纵坐标 $f_2^*$ 是约束问题(5.19)、(5.20)的最优目标值，且 $f_1^* = \epsilon_1$，点 $A$ 在可行集 $F$ 的东北角边界 $N$ 上，所以点 $A$ 是原多目标规划的非劣解.

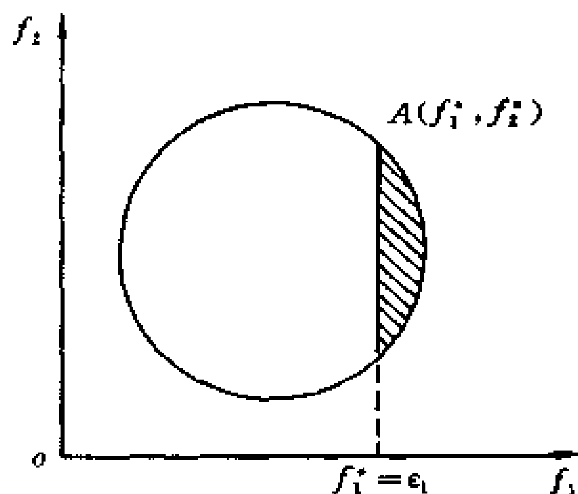



图 5.12

约束法的计算步骤

(1) 建立支付表

分别求单目标规划

$$\max \quad f_j(\boldsymbol{x}), \tag{5.21}$$

$$\text{s.t.} \quad \boldsymbol{x} \in X \tag{5.22}$$

的最优点 $\boldsymbol{x}^j=(x_1^j,\cdots,x_n^j)$，$j=1,\cdots,p$，建立**支付表**(pay-off table)

**表 5.1**

| | $f_1$ | $f_2$ | $\cdots$ | $f_p$ |
|---|---|---|---|---|
| | $f_1(\boldsymbol{x}^1)$ | $f_2(\boldsymbol{x}^1)$ | $\cdots$ | $f_p(\boldsymbol{x}^1)$ |
| | $f_1(\boldsymbol{x}^2)$ | $f_2(\boldsymbol{x}^2)$ | $\cdots$ | $f_p(\boldsymbol{x}^2)$ |
| | $\vdots$ | $\vdots$ | | $\vdots$ |
| | $f_1(\boldsymbol{x}^p)$ | $f_2(\boldsymbol{x}^p)$ | $\cdots$ | $f_p(\boldsymbol{x}^p)$ |

记 $M_j=f_j(\boldsymbol{x}^j)=\max\limits_{\text{s.t. }\boldsymbol{x}\in X} f_j(\boldsymbol{x})$，$m_j=\min\limits_{1\leqslant i\leqslant p} f_j(\boldsymbol{x}^i)$为各列的极小值，$j=1,\cdots,p$. 设

$$\varepsilon_j = m_j + \frac{t}{r-1}(M_j - m_j),\ t = 0,1,\cdots,r-1,$$

$$j = 1,\cdots,p,j \neq k \tag{5.23}$$

其中 $r$ 为给定的正整数. 选定主目标 $f_k(\boldsymbol{x})$.

(2) 令 $\varepsilon_j$ 在(5.23)规定的范围内取值，组合成不同的下界 $\boldsymbol{\varepsilon}_k=(\varepsilon_1,\cdots,\varepsilon_{k-1},\varepsilon_{k+1},\cdots,\varepsilon_p)$，对每个 $\boldsymbol{\varepsilon}_k$ 求约束问题 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$的最优解. 因为有 $r^{p-1}$个不同的 $\boldsymbol{\varepsilon}_k$，有 $r^{p-1}$个 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$问题，解这些问题可以得到许多非劣解，这些非劣解构成一个近似的非劣集.

**例 5.3.7** 用约束法求解双目标规划(5.12)，(5.13)

**解** (1) 求支付表 求出$\max\limits_{\text{s.t. }(x_1,x_2)\in X} f_1(x_1,x_2)$ 的极大点 $\boldsymbol{x}^1=(5,0)$，$f_1(\boldsymbol{x}^1)=10$，$f_2(\boldsymbol{x}^1)=-5$；再求出$\max\limits_{\text{s.t. }(x_1,x_2)\in X} f_2(x_1,x_2)$ 的极大点 $\boldsymbol{x}^2=(1,5)$，$f_1(\boldsymbol{x}^2)=-3$，$f_2(\boldsymbol{x}^2)=14$，支付表为

| | $f_1$ | $f_2$ |
|---|---|---|
| | 10 | $-5$ |
| | $-3$ | 14 |

选定 $f_2(x_1,x_2)$为主目标. 考虑约束问题中 $f_1(x_1,x_2)$的下界. 支付表中第一列的极大值为 $M_1=10$，极小值为 $m_1=-3$. 将区间$[m_1,$

$M_1]=[-3,10]$ $r-1$ 等分(令 $r=14$),得到分点为

$$\varepsilon_1 = m_1 + \frac{t}{r-1}(M_1 - m_1)$$
$$= -3, -2, -1, \cdots, 8, 9, 10, t = 0, 1, \cdots, r-1.$$

(2) 求约束问题的最优解

$$\begin{aligned} \max \quad & f_2(x_1,x_2) = -x_1 + 3x_2; \\ \text{s.t.} \quad & f_1(x_1,x_2) = 2x_1 - x_2 \geqslant \varepsilon_1, \\ & -x_1 + x_2 \leqslant 4, \\ & x_1 + x_2 \leqslant 9, \\ & 0 \leqslant x_1, x_2 \leqslant 5. \end{aligned}$$

将上述 $\varepsilon_1$ 的 14 个值分别代入上式. 解 14 个单目标线性规划,求出它们的最优解,这些最优解都是双目标规划(5.12)、(5.13)的非劣解. 现将这些非劣解列表如下:

**表 5.2**

| $\varepsilon_1$ | $x_1^*$ | $x_2^*$ | $f_1^*$ | $f_2^*$ | 在图中的位置 |
|---|---|---|---|---|---|
| −3 | 1 | 5 | −3 | 14 | $C$ 点 |
| −2 | 1.5 | 5 | −2 | 13.5 | $\overline{CD}$上的点 |
| −1 | 2 | 5 | −1 | 13 | $\overline{CD}$上的点 |
| 0 | 2.5 | 5 | 0 | 12.5 | $\overline{CD}$上的点 |
| 1 | 3 | 5 | 1 | 12 | $\overline{CD}$上的点 |
| 2 | 3.5 | 5 | 2 | 11.5 | $\overline{CD}$上的点 |
| 3 | 4 | 5 | 3 | 11 | $D$ 点 |
| 4 | 4.333 | 4.666 | 4 | 9.666 | $\overline{DE}$上的点 |
| 5 | 4.666 | 4.333 | 5 | 8.333 | $\overline{DE}$上的点 |
| 6 | 5 | 4 | 6 | 7 | $E$ 点 |
| 7 | 5 | 3 | 7 | 4 | $\overline{EF}$上的点 |
| 8 | 5 | 2 | 8 | 1 | $\overline{EF}$上的点 |
| 9 | 5 | 1 | 9 | −2 | $\overline{EF}$上的点 |
| 10 | 5 | 0 | 10 | −5 | $F$ 点 |

从这张表可以看出，与上述加权法（见例 5.3.4）相似，由约束法也可以近似地得到该问题的非劣集（折线$\overline{CDEF}$）。参考图 5.10 与 5.11.

### 5.3.3 混合法

把上述加权法与约束法联合使用的方法叫**混合法**（hybrid method），混合法的基础是下述定理.

**定理 5.3.8** $x^*$ 是多目标规划（VOP）的非劣解的充要条件是：对于任意给定的一组 $w_i^0>0, i=1,\cdots,p$，存在一组实数 $\varepsilon_j (j=1,\cdots,p)$，使 $x^*$ 是

$$\begin{aligned}\max \quad & \sum_{i=1}^{p} w_i^0 f_i(x); \\ \text{s.t.} \quad & f_j(x) \geqslant \varepsilon_j, j=1,\cdots,p, \\ & x \in X\end{aligned}$$

的一个最优解.

像约束法那样，有规律地变动$(\varepsilon_1,\cdots,\varepsilon_p)$的值，求解一系列上述问题，可以得到近似的非劣集. 因为此定理的条件较弱，混合法比加权法或约束法方便.

## 5.4 寻求线性多目标规划非劣解的方法

### 5.4.1 线性多目标规划的单纯形法

**多目标单纯形法**（multiobjective simplex method）是线性规划单纯形法在线性多目标规划上的推广. 多目标规划中的目标函数与约束函数都是线性函数时称为线性多目标规划. 在线性规划中用单纯形法求极点最优解，而在线性多目标规划中用单纯形法求非劣极点解（noninferior extreme solution）.

设线性多目标规划为

(VLP) $\max \quad \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x};$ (5.24)

s.t. $\boldsymbol{x} \in X = \{\boldsymbol{x} \in E^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqslant \boldsymbol{0}\}.$ (5.25)

其中

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_p(\boldsymbol{x}))^{\mathrm{T}},$$

$$\boldsymbol{b} = (b_1, \cdots, b_m)^{\mathrm{T}},$$

$\boldsymbol{C}$ 是 $p \times n$ 矩阵，$\boldsymbol{A}$ 是 $m \times n$ 矩阵，

且

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} c_{ij} x_j,\ i = 1, \cdots, p. \tag{5.26}$$

与单目标线性规划的单纯形法相似，也要建立单纯形表. 若已知 $\boldsymbol{x}$ 的一个基本可行解[*]为 $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{pmatrix}$，其中 $\boldsymbol{x}_B$ 为基变量部分，$\boldsymbol{x}_N$ 为非基变量部分，相应地 $\boldsymbol{c}, \boldsymbol{A}$ 也可以分解为 $\boldsymbol{c} = (\boldsymbol{c}_B, \boldsymbol{c}_N)$ 与 $\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N})$，则约束条件 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 可以写成 $\boldsymbol{B}\boldsymbol{x}_B + \boldsymbol{N}\boldsymbol{x}_N = \boldsymbol{b}$，由此式可以推出 $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{x}_N$，代入 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x}$，目标函数可以写成 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b} - (\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_N)\boldsymbol{x}_N$，因 $\boldsymbol{x}_N = \boldsymbol{0}$，则 $\boldsymbol{x}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$，$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$. 在极点可行解 $\boldsymbol{x}$ 处的向量形式单纯形表为

| | $\boldsymbol{x}_B$ | $\boldsymbol{x}_N$ | $\bar{\boldsymbol{b}}$ |
|---|---|---|---|
| $\boldsymbol{x}_B$ | $\boldsymbol{I}$ | $\boldsymbol{B}^{-1}\boldsymbol{N}$ | $\boldsymbol{B}^{-1}\boldsymbol{b}$ |
| $\boldsymbol{f}$ | $\boldsymbol{0}$ | $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{c}_N$ | $\boldsymbol{c}_B\boldsymbol{B}^{-1}\boldsymbol{b}$ |

在极点可行解 $\boldsymbol{x}$ 处的数字形式单纯形表为表 $T(\boldsymbol{x})$，见表 5.3。

[*] 基本可行解的定义与线性规划的基本可行解相同(见定义 1.1.6).

表 5.3

| | $x_1$ | $\cdots$ | $x_m$ | $x_{m+1}$ | $\cdots$ | $x_n$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $\cdots$ | 0 | $y_{1,m+1}$ | $\cdots$ | $y_{1,n}$ | $\bar{b}_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_m$ | 0 | $\cdots$ | 1 | $y_{m,m+1}$ | $\cdots$ | $y_{m,n}$ | $\bar{b}_m$ |
| $f_1$ | 0 | $\cdots$ | 0 | $\pi_{1,m+1}$ | $\cdots$ | $\pi_{1,n}$ | $f_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $f_p$ | 0 | $\cdots$ | 0 | $\pi_{p,m+1}$ | $\cdots$ | $\pi_{p,n}$ | $f_p$ |

表中 $x_1,\cdots,x_m$ 是基变量,$x_{m+1},\cdots,x_n$ 是非基变量,$\bar{b}$ 为右端项.表的前 $m$ 行的含意与线形规划单纯形表相同(参看表 1.1),表的后面 $p$ 行是检验数行,每行是一个目标函数的检验数.检验数的定义与线性规划中的相同.

$$\pi_{lj}=\begin{cases}0, & j=1,\cdots,m,\\ \sum_{k=1}^{m}c_{lk}y_{kj}-c_{lj}, & j=m+1,\cdots,n,\end{cases}\quad l=1,\cdots,p.$$

关于线性多目标规划有下述定理.

**定理 5.4.1** 设 $\bar{x}$ 是(VLP)的一个基本可行解,且相应的单纯形表中第 $l$ 行检验数 $\pi_{lj}\geqslant 0,j=1,\cdots,n$,则 $\bar{x}$ 是单目标规划

$$\begin{aligned}&\max \quad f_l(x);\\ &\text{s.t.}\quad x\in X\end{aligned}$$

的最优解,又若 $\pi_{lj}>0,j=m+1,\cdots,n$,则此最优解是唯一的,且 $\bar{x}$ 是(VLP)的非劣解.

**定理 5.4.2** 设 $\bar{x}$ 是(VLP)的一个非退化的(即所有的基变量$>0$)基本可行解,若有一个非基列 $j$ 的所有检验数 $\pi_{lj}\leqslant 0,l=1,\cdots,p$,且至少有一个 $\pi_{lj}<0$,则 $\bar{x}$ 不是非劣解.

**定理 5.4.3** 若 $\bar{x}$ 是(VLP)的一个基本可行解,且辅助问题

$$\text{(SP)} \quad \max \quad V = \sum_{i=1}^{p} \delta_i; \tag{5.27}$$

$$\left.\begin{aligned} &\text{s.t.} \quad f_i(\boldsymbol{x}) - \delta_i = f_i(\overline{\boldsymbol{x}}), i = 1, \cdots, p, \\ &\boldsymbol{x} \in X, \\ &\delta_i \geqslant 0, i = 1, \cdots, p \end{aligned}\right\} \tag{5.28}$$

的最优值 $\max V = 0$，则 $\overline{\boldsymbol{x}}$ 是(VLP)的非劣解；若 $\max V > 0$，则 $\overline{\boldsymbol{x}}$ 不是(VLP)的非劣解.

**定理 5.4.4** 设 $\overline{\boldsymbol{x}}$ 是(VLP)的一个非退化的基本可行解，若有两个不同的非基向量 $\boldsymbol{a}_j$ 与 $\boldsymbol{a}_q$，$Q_j$ 是 $x_j$ 进基后所取的值，$Q_q$ 是 $x_q$ 进基后所取的值. 若 $\theta_j \pi_{ij} \leqslant \theta_q \pi_{iq} (i = 1, \cdots, p)$，且其中至少有一个严格不等式成立，则 $x_j$ 进基优于 $x_q$ 进基.

**定理 5.4.5** 设 $\overline{\boldsymbol{x}}$ 是(VLP)的一个非退化的基本可行解，若第 $j$ 列是非基列，此列的所有检验数 $\pi_{ij} \geqslant 0 (i = 1, \cdots, p)$，且其中至少有一个严格不等式成立，则 $x_j$ 进基将不会得到非劣解.

线性多目标规划单纯形算法的计算步骤：

应用上述定理，可以构造求线性多目标规划非劣极点解集的单纯形法，其步骤由图 5.13 给出.

**例 5.4.6** 用单纯形法求下述问题的所有非劣极点解.

$$\max \quad \boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + x_2 \end{pmatrix};$$

$$\begin{aligned} \text{s.t.} \quad & \boldsymbol{x} \in X, \\ & X = \{\boldsymbol{x} \in E^2 \mid x_1 + x_2 \leqslant 8, 2x_1 + x_2 \leqslant 11, \\ & \qquad 0 \leqslant x_1 \leqslant 5, 0 \leqslant x_2 \leqslant 6\}. \end{aligned}$$

**解** 加上松弛变量后，约束条件可写成

$$\begin{aligned} x_1 + x_2 + x_3 \qquad\qquad &= 8, \\ 2x_1 + x_2 \quad + x_4 \qquad &= 11, \\ x_1 \qquad\qquad + x_5 \quad &= 5, \\ x_2 \qquad\qquad + x_6 &= 6, \\ x_i \geqslant 0, \quad i = 1, \cdots, 6&. \end{aligned}$$

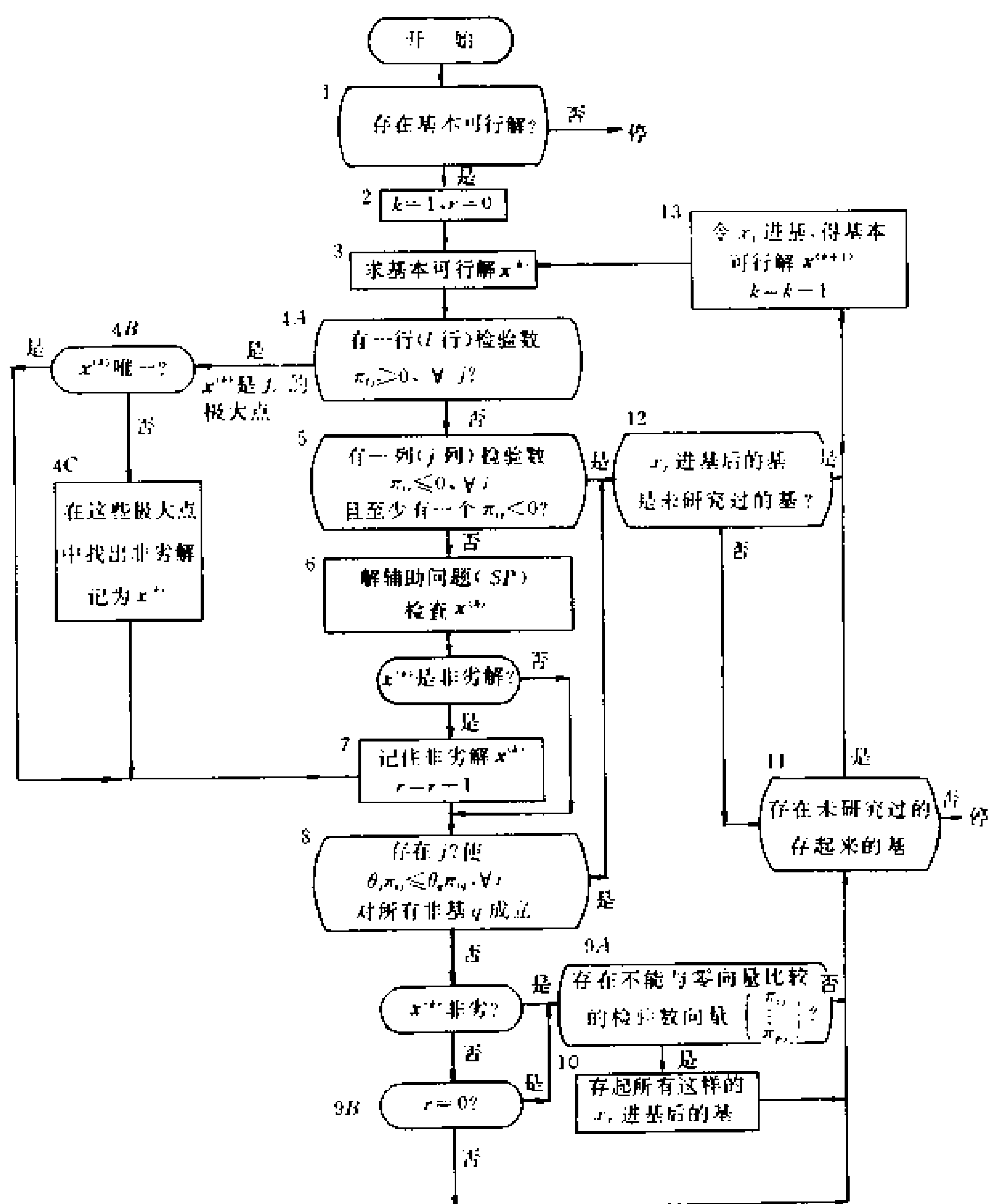


图 5.13

首先找到一个初始基本可行解 $x^{(1)}=(5,1,2,0,0,5)$，对应的初始

单纯形表为

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_3$ | 0 | 0 | 1 | $-1$ | 1 | 0 | 2 |
| $x_2$ | 0 | 1 | 0 | 1 | $-2$ | 0 | 1 |
| $x_1$ | 1 | 0 | 0 | 0 | 1 | 0 | 5 |
| $x_6$ | 0 | 0 | 0 | $-1$ | 2 | 1 | 5 |
| $f_1$ | 0 | 0 | 0 | 2 | $-3$ | 0 | 7 |
| $f_2$ | 0 | 0 | 0 | 1 | 1 | 0 | 16 |

(5.29)

在解法的第 4A 步(图 5.13),$f_2$ 的非基变量检验数$(\pi_{24},\pi_{25})=(1,1)>\mathbf{0}$,$\boldsymbol{x}^{(1)}$是 $f_2$ 的唯一极大点,因而是多目标规划的非劣极点解.在第 9A 步,$x_5$ 可以进基,在第 13 步 $x_5$ 进基,得基本可行解 $\boldsymbol{x}^{(2)}=(3,5,0,0,2,1)$,相应的单纯形表为

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_5$ | 0 | 0 | 1 | $-1$ | 1 | 0 | 2 |
| $x_2$ | 0 | 1 | 2 | $-1$ | 0 | 0 | 5 |
| $x_1$ | 1 | 0 | $-1$ | 1 | 0 | 0 | 3 |
| $x_6$ | 0 | 0 | $-2$ | 1 | 0 | 1 | 1 |
| $f_1$ | 0 | 0 | 3 | $-1$ | 0 | 0 | 13 |
| $f_2$ | 0 | 0 | $-1$ | 2 | 0 | 0 | 14 |

(5.30)

第 6 步解辅助问题:

$$\begin{aligned}\max\quad & V=\delta_1+\delta_2;\\ \text{s.t.}\quad & x_1+2x_2-\delta_1=13,\\ & 3x_1+x_2-\delta_2=14,\\ & \boldsymbol{x}\in X,\\ & \delta_1,\delta_2\geqslant 0.\end{aligned}$$

得到 $\max V=0$,因此 $\boldsymbol{x}^{(2)}$是非劣极点解.在第 9A 步与第 11 步,$x_4$ 可以进基,且进基后的新基是未研究过的.在第 13 步,令 $x_4$ 进基,基本可行解为 $\boldsymbol{x}^{(3)}=(2,6,0,1,3,0)$,单纯形表为

|  | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_5$ | 0 | 0 | −1 | 0 | 1 | 1 | 3 |
| $x_3$ | 0 | 1 | 0 | 0 | 0 | 1 | 6 |
| $x_1$ | 1 | 0 | 1 | 0 | 0 | −1 | 2 |
| $x_4$ | 0 | 0 | −2 | 1 | 0 | 1 | 1 |
| $f_1$ | 0 | 0 | 1 | 0 | 0 | 1 | 14 |
| $f_2$ | 0 | 0 | 3 | 0 | 0 | −2 | 12 |

因为 $f_1$ 的非基变量检验数$(\pi_{13},\pi_{16})=(1,1)>\mathbf{0}$，$\boldsymbol{x}^{(3)}$是 $f_1$ 的唯一极大点，因而是非劣极点解. 直至计算终结，得不到其它非劣极点解. 已经得到的三个非劣极点解 $\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}$与 $\boldsymbol{x}^{(3)}$分别在图 5.14($a$)的$(x_1,x_2)$平面及图 5.14($b$)的$(f_1,f_2)$平面上用 $B$,$C$,$D$ 三点表示. 从图可以看出，折线$\overline{BCD}$是非劣集. $B$,$C$,$D$ 三点是全部非劣极点解.

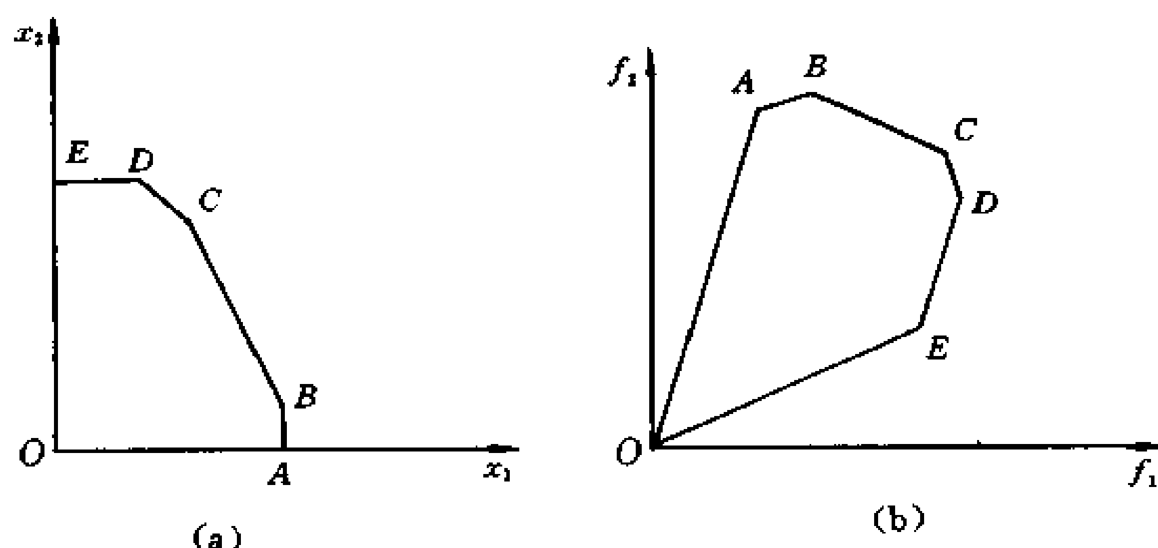



图 5.14

### 5.4.2 寻求线性多目标规划相邻非劣极点解的方法

下面介绍求一个非劣极点解的所有相邻非劣极点解的方法.

这个方法对以后要讲的交互法是有用的.设已知线性多目标规划为(5.24)、(5.25),即

$$(\text{VLP})\quad \max \quad \boldsymbol{cx};$$

$$\text{s.t.}\quad \boldsymbol{x} \in X = \left\{\boldsymbol{x} \in E^n \mid \sum_{j=1}^{n} a_{ij}x_j = b_i, i = 1,\cdots,m,\right.$$

$$\left. x_j \geqslant 0, j = 1,\cdots,n\right\}.$$

其中 $c$ 是 $p\times n$ 矩阵

**定义 5.4.7** 设 $\boldsymbol{x}^{(0)}$是(VLP)的一个非劣极点解.满足下列两个条件的 $\boldsymbol{x}^{(1)}$称为 $\boldsymbol{x}^{(0)}$的相邻非劣极点解.

(1) $\boldsymbol{x}^{(1)}$是由 $\boldsymbol{x}^{(0)}$的一个非基变量 $x_l$ 进基后得到的另一个极点.

(2) 任何 $\boldsymbol{x}=\alpha\boldsymbol{x}^{(0)}+(1-\alpha)\boldsymbol{x}^{(1)}$, $0\leqslant\alpha\leqslant 1$,都是(VLP)的非劣解.

上述非基变量 $x_l$ 称为**有效变量**(efficient variable),$\boldsymbol{x}^{(0)}$的所有有效变量的集称为**有效变量集**(efficient variable set),记作 $J(x^{(0)})$.

**定理 5.4.8** 设 $\boldsymbol{x}^{(0)}$是(VLP)的非劣极点解,则 $\boldsymbol{x}^{(1)}$是 $\boldsymbol{x}^{(0)}$的相邻非劣极点解的充要条件是

(1) $\boldsymbol{x}^{(1)}$是由 $\boldsymbol{x}^{(0)}$的一个非基变量 $x_l$ 进基后得到的另一个极点.

(2) 存在 $\boldsymbol{w}=(w_1,\cdots,w_p)>\boldsymbol{0}$,使 $\boldsymbol{x}^{(0)}$与 $\boldsymbol{x}^{(1)}$都是加权问题

$$\text{P}(\boldsymbol{w})\qquad \max \quad \boldsymbol{wcx} = \sum_{i=1}^{p} w_i \sum_{j=1}^{n} c_{ij}x_j,$$

$$\text{s.t.}\quad \boldsymbol{x} \in X$$

的最优解.

上述定义 5.4.8 与定理 5.4.9 中的条件使用不方便,下面给出一个求相邻非劣极点解的实用方法.这里要用到(VLP)的单纯形表

$T(\boldsymbol{x}^{(0)})$(表 5.3)及加权问题 P(w)在 $\boldsymbol{x}^{(0)}$ 的单纯形表中的检验数

$$\pi_j = \sum_{i=1}^{p} w_i \pi_{ij}, \quad j = m+1, \cdots, n,$$

这里 $\pi_{ij}$ 是(VLP)单纯形表 $T(\boldsymbol{x}^{(0)})$ 中的检验数.

**定理 5.4.9** 已知 $\boldsymbol{x}^{(0)}$ 是(VLP)的非劣极点解,相应的单纯形表为 $T(\boldsymbol{x}^{(0)})$,若对于某个非基变量 $x_l$,问题

$$\delta_l = \min \epsilon_l;$$

$$\text{s.t.} \ -\sum_{i=1}^{p} \pi_{ij} + \epsilon_j = 0, j = m+1, \cdots, n,$$

$$\sum_{i=1}^{p} w_i = 1,$$

$$w_i \geqslant 0, i = 1, \cdots, p,$$

$$\epsilon_j \geqslant 0, j = m+1, \cdots, n.$$

有最优解 $(w_1, \cdots, w_p) > \mathbf{0}$,且最优值 $\delta_l = 0$,则 $x_l$ 进基后所得的极点 $\boldsymbol{x}^{(1)}$ 是 $\boldsymbol{x}^{(0)}$ 的相邻非劣极点解,$x_l$ 是有效变量.

用定理 5.4.10 提供的方法,逐个检查所有非基变量,可以得到 $\boldsymbol{x}^{(0)}$ 的所有相邻非劣极点解.

**例 5.4.10** 已知 $\boldsymbol{x}^{(1)} = [6 \quad 0 \quad 0]^{\mathrm{T}}$ 是

$$\max \quad \boldsymbol{cx} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 0 \\ 3 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}; \tag{5.31}$$

$$\text{s.t.} \quad \left.\begin{aligned} x_1 + x_2 \qquad &\leqslant 6, \\ x_2 \qquad &\leqslant 3, \\ x_1 + x_2 - x_3 &\geqslant 4, \\ x_1, x_2, x_3 &\geqslant 0. \end{aligned}\right\} \tag{5.32}$$

的一个非劣极点解,求与它相邻的所有非劣极点解.

**解** 对三个约束条件引进松弛变量 $x_4$、$x_5$ 与 $x_6$ 以后,可以得到在 $\boldsymbol{x}^{(1)}$ 处的单纯形表 5.4

表 5.4

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 1 | 0 | 1 | 0 | 0 | 6 |
| $x_5$ | 0 | 1 | 0 | 0 | 1 | 0 | 3 |
| $x_6$ | 0 | 0 | 1 | 1 | 0 | 1 | 2 |
| $f_1$ | 0 | 1 | −1 | 1 | 0 | 0 | 6 |
| $f_2$ | 0 | −1 | 0 | 2 | 0 | 0 | 12 |
| $f_3$ | 0 | 3 | 2 | 3 | 0 | 0 | 18 |

因为非基变量 $x_4$ 的检验数列$(1,2,3)^{\mathrm{T}}>\mathbf{0}$，按定理 5.4.5，$x_4$ 不能进基，只考虑 $x_2$ 与 $x_3$ 进基的可能性. 对于 $l=2$ 与 $l=3$ 分别求解

$$
\begin{aligned}
\delta_l &= \min \varepsilon_l; \\
\text{s.t.} \; -w_1 + w_2 - 3w_3 + \varepsilon_2 &= 0, \\
w_1 - 2w_3 + \varepsilon_3 &= 0, \\
-w_1 - 2w_2 - 3w_3 + \varepsilon_4 &= 0, \\
w_1 + w_2 + w_3 &= 1, \\
w_1, w_2, w_3, \varepsilon_2, \varepsilon_3, \varepsilon_4 &\geqslant 0.
\end{aligned}
$$

最优解是$(w_1,w_2,w_3,\varepsilon_2,\varepsilon_3,\varepsilon_4)=\left(\frac{2}{8},\frac{5}{8},\frac{1}{8},0,0,\frac{15}{8}\right)$，而且是 $l=2$ 与 $l=3$ 两个问题的最优解，并有 $\delta_2=0$，$\delta_3=0$，根据定理 5.4.10，$x_2$ 与 $x_3$ 都是有效变量，有效变量集是 $J(\boldsymbol{x}^{(1)})=\{2,3\}$，$x_2$ 与 $x_3$ 分别进基后得到的极点解 $\boldsymbol{x}^{(2)}=(3,3,0)^{\mathrm{T}}$ 与 $\boldsymbol{x}^{(3)}=(6,0,2)^{\mathrm{T}}$ 都是 $\boldsymbol{x}^{(1)}$的相邻非劣极点解.

可以验证 $\boldsymbol{x}^{(i)}$，$i=1,2,3$ 都是加权问题

$$
\begin{aligned}
\mathrm{P}(\boldsymbol{w}) \qquad \max F(\boldsymbol{x}) &= \sum_{i=1}^{3} w_i f_i(\boldsymbol{x}) \\
&= \frac{2}{8} f_1(\boldsymbol{x}) + \frac{5}{8} f_2(\boldsymbol{x}) + \frac{1}{8} f_3(\boldsymbol{x}) \\
&= \frac{1}{8}(15x_1 + 15x_2);
\end{aligned}
$$

$$\text{s.t.} \quad \boldsymbol{x} \in X$$

的最优解，最优函数值为 $F(\boldsymbol{x}^{(1)})=F(\boldsymbol{x}^{(2)})=F(\boldsymbol{x}^{(3)})=90/8$. 从图 5.15 也可以看出，$\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}$与 $\boldsymbol{x}^{(3)}$在 $F(\boldsymbol{x})$的等值面上.

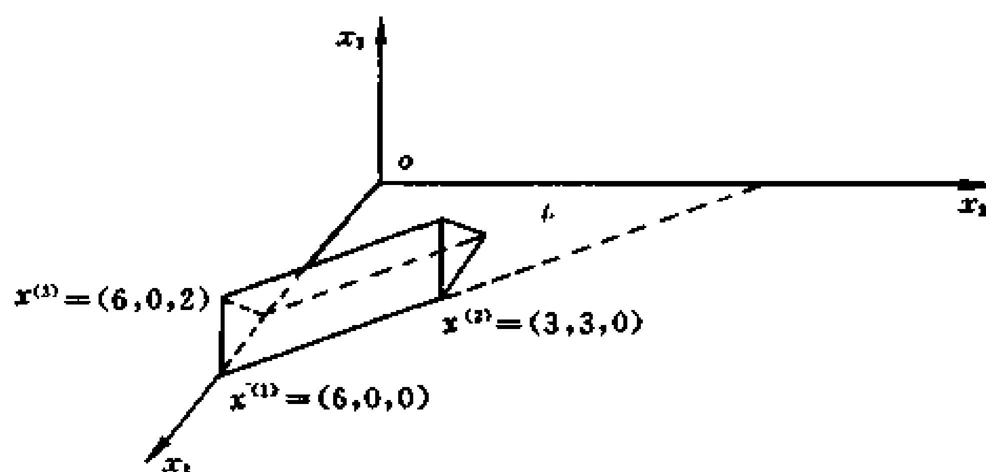



图 5.15

## 5.5 解多目标规划的交互法

**交互法**(interactive method)是一类需要在解题过程中与决策者对话的方法，通常通过人机对话来实现，因为计算量较小，有决策者参加，是比较有效的方法. 下面介绍几种交互法.

### 5.5.1 Geoffrion 方法

Geoffrion 方法是非线性规划中 Frank-Wolfe 方法(简称 F-W 方法)在多目标规划中的应用. 若线性约束多目标规划为

$$(\text{VLCP}) \qquad \max[f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})];$$

$$\text{s.t.} \sum_{j=1}^{n} a_{ij}x_j = b_i, i=1,\cdots,m,$$

$$x_j \geqslant 0, j=1,\cdots,n.$$

可行集 $X=\{x\in E^n \mid \sum_{j=1}^{n}a_{ij}x_j=b_i, i=1,\cdots,m, x_j\geqslant 0, j=1,\cdots,n\}$. 若已知(VLCP)决策者的效用函数 $U(f_1,\cdots,f_p)$(定义 5.2.9),只要求出非线性规划

$$\begin{aligned} \max \quad & U(f_1(x),\cdots,f_p(x)); \\ \text{s.t.} \quad & x\in X \end{aligned} \tag{5.33}$$

的最优解就得到最终解了. 但在很多实际问题中不知道 $U$ 的表达式,但可以得到边际替代率(见定义 5.2.14)的信息,也就是关于 $\partial U/\partial x_i$ 的信息,利用这些信息使问题得到解决. 此法的每一次迭代已知一个可行解 $x^{(k)}$,然后求解一个线性规划

$$\begin{aligned} \max \quad & \sum_{i=1}^{n}\left.\frac{\partial U}{\partial x_i}\right|_{x=x^{(k)}} y_i; \\ \text{s.t.} \quad & (y_1,\cdots,y_n)\in X. \end{aligned} \tag{5.34}$$

因为不知道$\left.\frac{\partial U}{\partial x_i}\right|_{x=x^{(k)}}$而有困难,但由(5.6)可知

$$\left.\frac{\dfrac{\partial U}{\partial x_j}}{\dfrac{\partial U}{\partial x_l}}\right|_{x=x^{(k)}}=m_{lj}^{(k)},\quad l,j=1,\cdots,p,\quad j\neq l.$$

这里 $m_{lj}^{(k)}$ 是边际替代率,对某些问题,决策者可以提供 $m_{lj}^{(k)}$ 的值,则(5.34)的目标函数可以简化

$$\begin{aligned} \sum_{i=1}^{n}\left.\frac{\partial U}{\partial x_i}\right|_{x=x^{(k)}} y_i &= \sum_{i=1}^{n}\left.\left(\sum_{j=1}^{p}\frac{\partial U}{\partial f_j}\frac{\partial f_j}{\partial x_i}\right)\right|_{x=x^{(k)}} y_i \\ &= \left(\frac{\partial U}{\partial f_l}\right)_{x=x^{(k)}}\cdot\sum_{i=1}^{n}\left[\sum_{j=1}^{p}\left(\frac{\dfrac{\partial U}{\partial f_j}}{\dfrac{\partial U}{\partial f_l}}\right)_{x=x^{(k)}}\frac{\partial f_j(x^{(k)})}{\partial x_i}\right] y_i \\ &= \left(\frac{\partial U}{\partial f_l}\right)_{x=x^{(k)}}\sum_{i=1}^{n}\left(\sum_{j=1}^{p}m_{lj}^{(k)}\frac{\partial f_j(x^{(k)})}{\partial x_i}\right) y_i. \end{aligned}$$

这里 $l$ 是 $1,\cdots,p$ 中任一数,为了方便,不妨设 $l=1$,在忽略常数因

子$\left(\frac{\partial U}{\partial f_1}\right)_{x=x^{(k)}}$以后，(5.34)可以改写为

$$\max \sum_{i=1}^{n}\left(\sum_{j=1}^{p} m_{1j}^{(k)} \frac{\partial f_j(\boldsymbol{x}^{(k)})}{\partial x_i}\right) y_i; \tag{5.35}$$

$$\text{s.t.} \quad (y_1, \cdots, y_n) \in X. \tag{5.36}$$

这里 $y_i$ 的系数是已知数，这是一个可以求解的线性规划问题.

Geoffrion 方法的计算步骤：

(1) 求出(VLCP)的一个初始可行点 $\boldsymbol{x}^{(1)} \in X$，令 $k=1$.

(2) 求搜索方向　在第 $k$ 次迭代，将已求出的迭代点 $\boldsymbol{x}^{(k)}$ 的目标函数值 $f_j(\boldsymbol{x}^{(k)})$ 提供给决策者，并要求决策者考虑，在这种水平的目标下，愿意用减少多少单位的 $f_1$ 值以换取 $f_j$ 增加 1 个单位，这个值就是边际替代率 $m_{1j}^{(k)} = -\frac{\partial f_1}{\partial f_j}$，请决策者估出所有的 $m_{1j}^{(k)}$，$j=2,3,\cdots,p$，并规定 $m_{11}^{(k)}=1$，然后求出线性规划(5.35)、(5.36)的最优解 $\boldsymbol{y}^{(k)}$. 以 $\boldsymbol{d}^{(k)} = \boldsymbol{y}^{(k)} - \boldsymbol{x}^{(k)}$ 作为搜索方向.

(3) 一维搜索　以 $\lambda$ 作为步长因子，$0 \leqslant \lambda \leqslant 1$，将$[0,1]$ $l$ 等分(一般令 $l=10$)，把 $\lambda = 0, \frac{1}{l}, \frac{2}{l}, \cdots, 1$ 分别代入$(f_1(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)}), \cdots, f_p(\boldsymbol{x}^{(k)}+\lambda\boldsymbol{d}^{(k)}))$并得出 $l+1$ 组目标值，请决策者从中选出最满意的一组，记此组的 $\lambda$ 值为 $\lambda^{(k)}$，以 $\lambda^{(k)}$ 作为最优步长. 令

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda^{(k)}\boldsymbol{d}^{(k)}.$$

(4) 终止判断　若 $\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| < \varepsilon$，则认为 $\boldsymbol{x}^{(k)}$ 是最终解，停止计算；否则，令 $k=k+1$，转(2).

### 5.5.2　逐步法

逐步法(step method，简记为 STEM)是 1971 年 Benayoun 等人提出的解线性多目标规划的一种交互式迭代法. 此法是在某种范数下求距理想点最短的点. 在每次迭代中，分析者向决策者提供试验解及相应的目标函数值，请决策者指出哪一个目标值应该增

加，哪一个目标值可以减少．分析者根据决策者的意图，增添新的约束，求新的试验解，进入下一次迭代．这样逐步修改，直到求出使决策者满意的最终解为止．若目标函数的个数是 $p$，此法一般只进行 $p$ 次迭代．$p$ 次迭代以后仍没有找到满意的解，此法失败，应该改用其它方法．

**定义 5.5.1** 给定权向量 $\boldsymbol{w}=(w_1,\cdots,w_p)$，$w_i\geqslant 0$，$i=1,\cdots,p$，在 $l_\infty$ 范数下二点 $\boldsymbol{f}=(f_1,\cdots,f_p)$ 与 $\boldsymbol{M}=(m_1,\cdots,m_p)$ 之间的距离为

$$\|\boldsymbol{f}-\boldsymbol{M}\|_\infty^w=\max_{i=1,\cdots,p}\{w_i|f_i-M_i|\}.$$

设线性多目标规划为

$$\text{(VLP)}\quad \max\ [f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})];$$

$$\text{s.t.}\ \boldsymbol{x}\in X=\left\{\boldsymbol{x}\in E^n\,\middle|\,\sum_{j=1}^n a_{ij}x_j=b_i,i=1,\cdots,m,\ x_j\geqslant 0,j=1,\cdots,n\right\}.$$

其中

$$f_i(\boldsymbol{x})=\sum_{j=1}^n c_{ij}x_j,\ i=1,\cdots,p.$$

逐步法的计算步骤：

(1) 建立支付表 求

$$\max\ f_j(\boldsymbol{x});$$
$$\text{s.t.}\ \boldsymbol{x}\in X$$

的极大点 $\boldsymbol{x}^j$，$j=1,\cdots,p$．建立下述支付表

表 5.5

| | $f_1$ | $f_2$ | $\cdots$ | $f_p$ |
|---|---|---|---|---|
| | $f_1(\boldsymbol{x}^1)$ | $f_2(\boldsymbol{x}^1)$ | $\cdots$ | $f_p(\boldsymbol{x}^1)$ |
| | $f_1(\boldsymbol{x}^2)$ | $f_2(\boldsymbol{x}^2)$ | | $f_p(\boldsymbol{x}^2)$ |
| | $\vdots$ | $\vdots$ | | $\vdots$ |
| | $f_1(\boldsymbol{x}^p)$ | $f_2(\boldsymbol{x}^p)$ | $\cdots$ | $f_p(\boldsymbol{x}^p)$ |

各列的极大值为 $M_j=f_j(\boldsymbol{x}^j)=\max\limits_{\boldsymbol{x}\in X}f_j(\boldsymbol{x}),j=1,\cdots,p,\boldsymbol{M}=(M_1,\cdots,M_p)$ 为理想点. 又记各列的极小值为 $m_j=\min\limits_{i=1,\cdots,p}f_j(\boldsymbol{x}^i)$, $j=1,\cdots,p$. 令 $k=1$,转(2)

(2) 求第 $k$ 次迭代点　在第 $k$ 次迭代时,在缩小的可行集 $X^{(k)}$(当 $k=1$ 时 $X^{(1)}=X$,$k>1$ 时,$X^{(k)}$由后面(5.42)给出)中求距理想点最近的点,即求解

$$\begin{aligned}&\min_{\boldsymbol{x}}\max_{i=1,\cdots,p}\{w_j[M_j-f_j(\boldsymbol{x})]\};\\&\text{s.t. }\boldsymbol{x}\in X^{(k)}.\end{aligned}\tag{5.37}$$

这里

$$w_j=\frac{\alpha_j}{\sum\limits_{l=1}^{p}\alpha_l},j=1,\cdots,p.\tag{5.38}$$

其中

$$\alpha_j=\begin{cases}\dfrac{M_j-m_j}{M_j}\left[\sum\limits_{l=1}^{n}c_{jl}^2\right]^{-\frac{1}{2}}, & \text{当 } M_j>0,\\ \dfrac{m_j-M_j}{m_j}\left[\sum\limits_{l=1}^{n}c_{jl}^2\right]^{-\frac{1}{2}}, & \text{当 } M_j\leqslant 0,\end{cases}\quad j=1,\cdots,p.\tag{5.39}$$

$c_{jl}$是线性目标函数 $f_j(\boldsymbol{x})$的系数. 由(5.38)给出的 $w_j$ 叫权因子,$w_j\geqslant 0$,且 $\sum\limits_{j=1}^{p}w_j=1$. 问题(5.37)是一个极小-极大问题,引进新的辅助变量 $\lambda$ 以后,这个问题与下述线性规划等价:

$$\min\ \lambda;\tag{5.40}$$

$$\left.\begin{aligned}\text{s.t. }&w_j[M_j-f_j(\boldsymbol{x})]\leqslant\lambda,j=1,\cdots,p,\\&\boldsymbol{x}\in X^{(k)},\\&\lambda\geqslant 0.\end{aligned}\right\}\tag{5.41}$$

求出该问题的最优解 $\boldsymbol{x}^{(k)}$ 及 $\lambda^{(k)}$,计算出相应的目标函数值 $f_1(\boldsymbol{x}^{(k)}),\cdots,f_p(\boldsymbol{x}^{(k)})$.

(3) 与决策者对话

把目标函数值 $f_i(x^{(k)}), i=1,\cdots,p$ 提供给决策者. 1) 若决策者对这组目标值表示满意,则 $x^{(k)}$ 是最终解,停止计算;2) 若决策者对所有的目标值都不满意,则此法失败,停止计算;3) 若决策者对一部分目标值满意,对另一部分目标值不满意,则进一步要求决策者在满意的目标中选一个目标 $f_{j^*}$,并给出一个可以牺牲的量 $\Delta f_{j^*}$,意思是愿意让 $f_{j^*}(x^{(k)})$ 减少 $\Delta f_{j^*}$ 以换取其他不满意目标的增长. 为此,添加新约束,并令 $k=k+1$,规定

$$X^{(k)} = \{x \in X^{k-1} | f_{j^*}(x) \geqslant f_{j^*}(x^{(k-1)}) - \Delta f_{j^*};$$
$$f_j(x) \geqslant f_j(x^{(k-1)}), j = 1,\cdots,p, j \neq j^*\}. \quad (5.42)$$

令 $\alpha_{j^*}=0$. 若 $k>p$,停止计算,此法失败;否则,转(2).

**例 5.5.2** 用逐步法求解多目标规划(5.12),(5.13)

**解** (1) 建立支付表

求出 $\begin{matrix}\max f_1(x)\\ \text{s.t.}\ x\in X\end{matrix}$ 的最优解 $x^1=(5,0)$, $f_1(x^1)=10$, $f_2(x^1)=-5$,再求出 $\begin{matrix}\max f_2(x)\\ \text{s.t.}\ x\in X\end{matrix}$ 的最优解 $x^2=(1,5)$, $f_1(x^2)=-3$, $f_2(x^2)=14$. 支付表是

| | $f_1$ | $f_2$ |
|---|---|---|
| | 10 | −5 |
| | −3 | 14 |

二列的最大值与最小值是 $M_1=10, m_1=-3, M_2=14, m_2=-5$.

(2) 求迭代点

按(5.38)与(5.39)计算出:

$$\alpha_1 = 0.581, \alpha_2 = 0.429, w_1 = 0.575, w_2 = 0.425.$$

解下述线性规划

$$\min \quad \lambda;$$

$$\text{s.t.} \quad 0.575(10-2x_1+x_2)-\lambda \leqslant 0,$$

$$0.425(14+x_1-3x_2)-\lambda \leqslant 0,$$

$$x \in X^{(1)} = X = \{x \in E^2 \mid -x_1+x_2 \leqslant 4, \; x_1+x_2 \leqslant 9, 0 \leqslant x_1, x_2 \leqslant 5\},$$

$$\lambda \geqslant 0.$$

得到最优解 $x^{(1)}=(4.80, 4.20)$，$(f_1(x^{(1)}), f_2(x^{(1)}))=(5.40, 7.80)$。

(3) 与决策者对话

决策者对第一个目标值 $f_1(x^{(1)})=5.40$ 不满意，认为太小了，对 $f_2(x^{(1)})$ 满意，并且认为即使减少 0.5 个单位也可以，希望通过减少 $f_2$ 的值来换取提高 $f_1$ 的值，令 $j^*=2$，$\Delta f_{j^*}=\Delta f_2=0.5$，按规定 $\alpha_{j^*}=\alpha_2=0$。

$$X^{(2)} = \{x \in X^{(1)} \mid f_1(x) = 2x_1 - x_2 \geqslant 5.40, \; f_2(x) = -x_1 + 3x_2 \geqslant 7.30\},$$

$k=k+2$，转(2)。

(2) 求第二次迭代点

求出 $\alpha_1=0.581$，已知 $\alpha_2=0$，求出 $w_1=1$，$w_2=0$。求解线性规划

$$\min \quad \lambda;$$

$$\text{s.t.} \quad 1\cdot(10-2x_1+x_2)-\lambda \leqslant 0,$$

$$(x_1, x_2) \in X^{(2)},$$

$$\lambda \geqslant 0.$$

得到最优解 $x^{(2)}=(4.925, 4.075)$，$(f_1(x^{(2)}), f_2(x^{(2)}))=(5.775, 7.30)$

(3) 与决策者对话

将 $x^{(2)}$ 与目标值 $f_1(x^{(2)})$ 与 $f_2(x^{(2)})$ 提供给决策者，决策者表

示满意，因此 $\boldsymbol{x}^{(2)}$ 是最终解，停止计算.

### 5.5.3 Zionts-Wallenius 方法

这是 Zionts 与 Wallenius 在 1976 年提出的一种交互法，主要用于解线性多目标规划.

设线性多目标规划为

$$\text{(VLP)} \quad \max \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{c}\boldsymbol{x}; \tag{5.43}$$

$$\text{s.t.} \quad \boldsymbol{x} \in X = \left\{ \boldsymbol{x} \in E^n \mid \sum_{j=1}^{n} a_{ij}x_j = b_i, i = 1,\cdots,m, \right.$$

$$\left. x_j \geqslant 0, j = 1,\cdots,n \right\}. \tag{5.44}$$

其中

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x}))^{\mathrm{T}},$$

$$\boldsymbol{c} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{p1} & \cdots & c_{pn} \end{bmatrix}.$$

显然有

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} c_{ij}x_j, i = 1,\cdots,p.$$

假设决策者有(隐含的)线性的效用函数

$$U(f_1,\cdots,f_p) = \sum_{i=1}^{p} w_i f_i(\boldsymbol{x}), \tag{5.45}$$

其中 $w_i > 0, i = 1,\cdots,p$，且 $\sum_{i=1}^{p} w_i = 1$，这里所有 $w_i$ 是未知的权因子. 求此问题的最终解实质上就是求

$$\begin{aligned} &\max U(f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})) = \sum_{i=1}^{p} w_i f_i(\boldsymbol{x}); \\ &\text{s.t.} \quad \boldsymbol{x} \in X \end{aligned} \tag{5.46}$$

的最优解.这个方法的特点是,通过与决策者对话,了解关于效用函数的信息,逐步找出合适的权因子 $w_i(i=1,\cdots,p)$. 每次迭代,求一个非劣极点解 $\boldsymbol{x}^{(k)}$ 及相应的有效变量(定义 5.4.8)$x_l$,决策者是否欢迎 $x_l$ 进基,可以由 $\Delta U>0$ 或 $\Delta U<0$ 表示,这里 $\Delta U$ 是当 $x_l$ 增加 1 个单位时 $U$ 的增量, $\Delta U=-\sum_{i=1}^{p}\pi_{il}w_i$, 其中 $\pi_{il}$ 是 $\boldsymbol{x}^{(k)}$ 的单纯形表 5.6 中的检验数. 从一系列 $\Delta U$ 的不等式可得到一系列 $w_i$ 的不等式,从而可以逐步缩小 $w_i$ 的变化范围.

**表 5.6 单纯形表 $T(\boldsymbol{x}^{(k)})$**

| | $x_1$ | $\cdots$ | $x_m$ | $x_{m+1}$ | $\cdots$ | $x_n$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | $\cdots$ | 0 | $y_{1,m+1}$ | $\cdots$ | $y_{1n}$ | $\bar{b}_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_m$ | 0 | $\cdots$ | 1 | $y_{m,m+1}$ | $\cdots$ | $y_{mn}$ | $\bar{b}_m$ |
| $f_1$ | 0 | $\cdots$ | 0 | $\pi_{1,m+1}$ | $\cdots$ | $\pi_{1n}$ | $f_1$ |
| $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $f_p$ | 0 | $\cdots$ | 0 | $\pi_{p,m+1}$ | | $\pi_{pn}$ | $f_p$ |

Zionts-Wallenius 方法的计算步骤:

(1) 令 $\Lambda^{(1)}=\{(w_1,\cdots,w_p)\in E^p\,|\,w_i>0,i=1,\cdots,p,\ \sum_{i=1}^{p}w_i=1\}$,任选 $(w_1^{(1)},\cdots,w_p^{(1)})\in\Lambda^{(1)}$. 令 $k=1$.

(2) 用线性规划单纯形法求解加权问题

$$\max\ \sum_{j=1}^{p}w_j^{(k)}f_j(\boldsymbol{x});$$
$$\text{s.t.}\ \boldsymbol{x}\in X.$$

得最优解 $\boldsymbol{x}^{(k)}$,它是(VLP)的非劣极点解,相应的单纯形表是表 5.6 的 $T(\boldsymbol{x}^{(k)})$.

(3) 求有效变量集

用 5.4.2 中的方法求出 $\boldsymbol{x}^{(k)}$ 的所有有效变量及有效变量集

$J(\boldsymbol{x}^{(k)})$.

(4) 与决策者对话

每个有效变量 $x_l$ 在单纯形表 $T(\boldsymbol{x}^{(k)})$ 中有检验数 $\pi_{il}, i=1, \cdots, p$，向决策者提问："你是否愿意接受 $f_1(\boldsymbol{x}^{(k)})$减少 $\pi_{1l}$个单位，$f_2(\boldsymbol{x}^{(k)})$减少 $\pi_{2l}$个单位，…，$f_p(\boldsymbol{x}^{(k)})$减少 $\pi_{pl}$个单位？"决策者的回答有三个可能性："是"、"否"或"我不知道"。若对于所有的有效变量 $x_l \in J(\boldsymbol{x}^{(k)})$，决策者对上述问题的回答都是"否"，表示决策者不愿改动非劣极点解 $\boldsymbol{x}^{(k)}$，则 $\boldsymbol{x}^{(k)}$是最终解，停止计算；否则，分三种情况：

1) 对回答为"是"的 $x_l$，构造不等式

$$\sum_{i=1}^{p} \pi_{il} w_i \leqslant -\varepsilon;$$

2) 对回答为"否"的 $x_l$，构造不等式

$$\sum_{i=1}^{p} \pi_{il} w_i \geqslant \varepsilon;$$

3) 对回答为"我不知道"的 $x_l$，构造等式 $\sum_{i=1}^{p} \pi_{il} w_i = 0$，转(5).

(5) 求新的权因子

构造 $\Lambda^{(k+1)} \leqslant \Lambda^{(k)}$，$(w_1, \cdots, w_p) \in \Lambda^{(k+1)}$应满足下述条件：

$$\sum_{i=1}^{p} \pi_{il} w_i \leqslant -\varepsilon,$$

$$\sum_{i=1}^{p} \pi_{il} w_i \geqslant \varepsilon,$$

$$\sum_{i=1}^{p} \pi_{il} w_i = 0,$$

$$w_i \geqslant \varepsilon, i = 1, \cdots, p.$$

求一组$(w_1^{(k+1)}, \cdots, w_p^{(k+1)}) \in \Lambda^{(k+1)}$. 令 $k=k+1$，转(2).

**例 5.5.3** 用 Zionts-Wallenius 方法求解问题(5.31)，

(5.32).

$$\max \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ f_3(\boldsymbol{x}) \end{bmatrix} = \boldsymbol{cx} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 0 \\ 3 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix};$$

$$\text{s.t. } \boldsymbol{x} \in X = \{\boldsymbol{x} \in E^3 | x_1 + x_2 \leqslant 6, x_2 \leqslant 3,$$
$$x_1 + x_2 - x_3 \leqslant 4, x_1, x_2, x_3 \geqslant 0\}.$$

**解** 一般情况下决策者效用函数的表达式是不知道的，但本例为了说明问题，假设它是已知的，其表达式为

$$U(f_1, f_2, f_3) = 0.1f_1 + 0.8f_2 + 0.1f_3.$$

下面来求解.

(1) 选定一组初值 $w_i^{(1)} = \frac{1}{3}, i = 1,2,3$，令 $k = 1$.

(2) 解加权问题

$$\max \sum_{i=1}^{3} w_i^{(1)} f_i(\boldsymbol{x});$$
$$\text{s.t. } \boldsymbol{x} \in X.$$

得到非劣极点解 $\boldsymbol{x}^{(1)} = (6,0,0)^{\mathrm{T}}$，相应的单纯形表见上一节的表 5.4，此时 $(f_1, f_2, f_3) = (6,12,18)$.

(3) 例 5.4.10 已经求出 $\boldsymbol{x}^{(1)}$ 的有效变量为 $x_2$ 与 $x_3$.

(4) 与决策者对话　考虑 $x_2$ 进基，因为单纯形表中 $x_2$ 的检验数为 $\pi_{12} = 1, \pi_{22} = -1, \pi_{32} = 3$，问决策者：你愿意 $f_1$ 减少 1 个单位，$f_2$ 增加 1 个单位，$f_3$ 减少 3 个单位吗？决策者考虑 $\Delta U = 0.1(-1) + 0.8(1) + 0.1(-3) = 0.4 > 0$，效益增长，因此回答为“是”. 再考虑 $x_3$ 进基，因为 $\pi_{13} = -1, \pi_{23} = 0, \pi_{33} = 2$，问决策者：你愿意 $f_1$ 增加 1 个单位，$f_2$ 不动，$f_3$ 减少 3 个单位吗？因为 $\Delta U = 0.1(1) + 0.8(0) + 0.1(-3) = -0.2 < 0$，效益减少，回答“否”. 根据这些回答，构造不等式组

$$
\begin{cases}
w_1 - w_2 + 3w_3 \leqslant -\varepsilon, \\
-w_1 \qquad + 2w_3 \geqslant \varepsilon, \\
w_1 + w_2 + w_3 \quad = 1, \\
\qquad w_i \geqslant \varepsilon, i = 1,2,3.
\end{cases}
$$

这里令 $\varepsilon=0.01$.

(5) 求得上述不等式组的一个可行解$(w_1^{(2)}, w_2^{(2)}, w_3^{(2)}) = (0.1, 0.75, 0.1)$,令 $k=k+1=2$,转(2).

(2) 解

$$
\max \sum_{i=1}^{3} w_i^{(2)} f_i(\boldsymbol{x}),
$$
$$
\text{s.t. } \boldsymbol{x} \in X.
$$

得非劣极点解 $\boldsymbol{x}^{(2)}=[3 \quad 3 \quad 0]^{\mathrm{T}}$,对应的多目标规划单纯形表为

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $\boldsymbol{b}$ |
|---|---|---|---|---|---|---|---|
| $x_1$ | 1 | 0 | 0 | 1 | −1 | 0 | 3 |
| $x_2$ | 0 | 1 | 0 | 0 | 1 | 0 | 3 |
| $x_3$ | 0 | 0 | 1 | 1 | 0 | 1 | 2 |
| $f_1$ | 0 | 0 | −1 | 1 | −1 | 0 | 3 |
| $f_2$ | 0 | 0 | 0 | 2 | 1 | 0 | 15 |
| $f_3$ | 0 | 0 | 2 | 3 | −3 | 0 | 9 |

(3) 求有效变量 $x_4$ 的检验数都大于零,由定理 5.4.5 $x_4$ 不能进基,$x_5$ 进基后将得到已考虑过的极点 $\boldsymbol{x}^{(1)}$,应放弃.经过计算,得到 $x_3$ 是有效变量.

(4) 与决策者对话,因为 $x_3$ 的检验数列为$(-1,0,2)^{\mathrm{T}}$,问:你愿不愿意 $f_1$ 增加 1 个单位,$f_2$ 不动,$f_3$ 减少 2 个单位.因为

$$
\Delta U = 0.1(1) + 0.8(0) + 0.1(-2) = -0.1 < 0,
$$

决策者回答“否”,这样最终解是 $\boldsymbol{x}^{(2)}$.

### 5.5.4 代替价值交换法

代替价值交换法(surrogate worth trade-off method)是Haimes等人提出的一种交互法. 此法在与决策者对话时,要求他对所提供的非劣解及相应的交换率(Lagrange 乘子)做出评价,依据这些评价,再求出新的非劣解,直到决策者满意为止. 先介绍预备知识:

**1 Lagrange 乘子的一些性质**

设多目标规划问题为

$$\text{(VOP)}\quad \max\ [f_1(x),\cdots,f_p(x)],$$
$$\text{s.t. } x\in X=\{x\in E^n \mid g_i(x)\leqslant 0, i=1,\cdots,m\}.$$

对于选定的 $k$,(VOP)的约束问题是

$$\mathrm{P}_k(\varepsilon_k)\qquad \max\ f_k(x);\tag{5.47}$$

$$\left.\begin{aligned}\text{s.t. } & f_j(x)\geqslant \varepsilon_j, j=1,\cdots,p, j\neq k,\\ & g_i(x)\leqslant 0, i=1,\cdots,m.\end{aligned}\right\}\tag{5.48}$$

这里 $\varepsilon_k=(\varepsilon_1,\cdots,\varepsilon_{k-1},\varepsilon_{k+1},\cdots,\varepsilon_p)$,这些 $\varepsilon_j$ 是给定的下界. 对于给定的 $\varepsilon_k$,可以求出 $P_k(\varepsilon_k)$的最优解 $\bar{x}$ 及相应的 Lagrange 乘子 $\pi_{kj}$与 $\mu_i$ 满足最优性条件

$$\left.\begin{aligned}&\nabla f_k(\bar{x})+\sum_{\substack{j=1\\ j\neq k}}^{p}\pi_{kj}\ \nabla f_j(\bar{x})-\sum_{i=1}^{m}\mu_i\ \nabla g_i(\bar{x})=0,\\ &\pi_{kj}[f_j(\bar{x})-\varepsilon_j]=0, \pi_{kj}\geqslant 0, j=1,\cdots,p, j\neq k,\\ &\mu_i g_i(\bar{x})=0,\quad \mu_i\geqslant 0, i=1,\cdots,m.\end{aligned}\right\}\tag{5.49}$$

**定义 5.5.4** 若存在$\delta>0$,使 $x^*$是(VOP)在 $X\cap N(x^*,\delta)=\{x\in E^n \mid \|x^*-x\|<\delta\}$内的非劣解(即不存在另一个 $x\in X\cap N(x^*,\delta)$使 $f_j(x)\geqslant f_j(x^*)$, $j=1,\cdots,m$ 成立),则称 $x^*$ 是(VOP)的一个**局部非劣解**(local noninferior solution). 局部非劣解的集

叫**局部非劣集**.

**定理 5.5.5** 若问题(VOP)中所有 $f_j$ 与 $g_i$ 都是二阶连续可微函数,$\bar{x}$ 是约束问题 $P_k(\varepsilon_k)$的最优解,$\pi_{kj}$与 $\mu_i$是相应的 Lagrange 乘子,他们满足最优性条件(5.49),假设

(1) $\bar{x}$ 是 $P_k(\varepsilon_k)$的可行集的正则点(定义 5.2.16);

(2) $\bar{x}$ 满足 $P_k(\varepsilon_k)$的二阶最优性充分条件;

(3) 若 $g_i(\bar{x})=0$,则 $\mu_i>0$,$i=1,\cdots,m$;

(4) $f_j(\bar{x})=\varepsilon_j$,且 $\pi_{kj}>0$,$j=1,\cdots,p$,$j\neq k$.

则在目标空间中的点$(\bar{f}_1,\cdots,\bar{f}_p)=(f_1(\bar{x}),\cdots,f_p(\bar{x}))$的一个邻域上,(VOP)的局部非劣集可以用函数 $f_k=f_k(f_1,\cdots,f_{k-1},f_{k+1},\cdots,f_p)$表示,且

$$\pi_{kj}=-\frac{\partial f_k(f_1,\cdots,f_{k-1},f_{k+1},\cdots,f_p)}{\partial f_j},j=1,\cdots,p,j\neq k.$$

上式表示,Lagrange 乘子 $\pi_{kj}$是目标 $f_k$ 对目标 $f_j$ 在局部非劣集上的交换率(定义 5.2.13),即在局部非劣集上,当其他目标值不动时,$f_k$ 对 $f_j$ 的变化率是$-\pi_{kj}$.

**2 代替价值函数**

已知一个非劣解 $\bar{x}$,目标函数值 $f_1(\bar{x}),\cdots,f_p(\bar{x})$及 Lagrange 乘子 $\pi_{kj}$以后,向决策者提问"在目标值为 $f_1(\bar{x}),\cdots,f_p(\bar{x})$时,你愿意 $f_j$ 增加 1 个单位,其他目标值不动,$f_k$ 减少 $\pi_{kj}$个单位吗?"决策者用给代替价值函数 $w_{kj}(\bar{x})$赋值来回答问题.

规定 $w_{kj}(\bar{x})$(简记为 $w_{kj}$)取值的范围是$-10$ 到$+10$ 之间的整数.取值的含意是:

(1) 若决策者同意上述交换,应给 $w_{kj}$赋正值,$w_{kj}$的值愈大表示愈赞成.

(2) 若决策者赞成反向的交换,即赞成用 $f_j$ 减少 1 个单位使 $f_k$ 增加 $\pi_{kj}$个单位,应给 $w_{kj}$赋负值,$|w_{kj}|$的值愈大表示愈赞成.

(3) 若对上述两种交换都不赞成，令 $w_{kj}=0$.

对这个非劣解 $\bar{x}$，决策者要对 $p-1$ 个 $w_{kj}, j=1,\cdots,p, j\neq k$，赋值. 若所有的 $w_{kj}(\bar{x})=0$，表示决策者对这组目标值满意，$\bar{x}$ 是最终解；否则，要设法构造出使所有 $w_{kj}$ 都等于 0 的非劣解来.

在 $p=2$ 时，代替价值函数 $w_{21}$ 有几何解释，在图 5.16 中，$N$ 是非劣集，$L_1, L_2$ 与 $L_3$ 是等效用曲线，$A$ 是非劣点，$\pi_{21}$ 是相应的拉格朗日乘子，$\pi_{21}$ 表示 $N$ 在 $A$ 点的负斜率. 可以看出，此斜率比 $L_1$ 在 $A$ 点的斜率小(指绝对值)，因此决策者愿意 $f_1$ 增加，令 $w_{21}>0$. 而在非劣点 $B$，非劣集 $N$ 的斜率比等效用曲线 $L_1$ 的斜率大，决策者愿意 $f_1$ 减少，令 $w_{21}<0$.

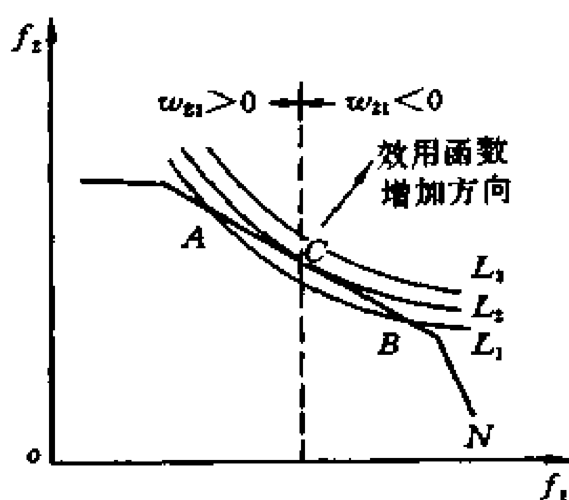



图 5.16

代替价值交换法的计算步骤：

求解多目标规划(VOP)的步骤是：

(1) 求目标函数的极小值与极大值　选一个 $f_k(x)$ 作为主要的目标函数，选定以后，就不再改变. 然后，分别求

$$\begin{array}{ll} \min f_j(x); & \max f_j(x); \\ \text{s.t. } x\in X. & \text{s.t. } x\in X \end{array}$$

与

的最优解，得到 $f_j(x)$ 的极小值 $m_j$ 与极大值 $M_j$，$j=1,\cdots,p,j\neq k$.

(2) 求非劣解　构造约束问题

$$
\mathrm{P}_k(\boldsymbol{\varepsilon}_k)\qquad \begin{aligned} &\max\ f_k(x);\\ &\text{s.t. } f_j(x)\geqslant \varepsilon_j,j=1,\cdots,p,j\neq k,\\ &\qquad g_i(x)\leqslant 0,i=1,\cdots,m. \end{aligned}
$$

这里 $\boldsymbol{\varepsilon}_k=(\varepsilon_1,\cdots,\varepsilon_{k-1},\varepsilon_{k+1},\cdots,\varepsilon_p)$，$\varepsilon_j=M_j-\bar{\varepsilon}_j$（给定 $\bar{\varepsilon}_j>0$），是选定的下界. 求 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$ 的最优解及相应的 Lagrange 乘子 $\pi_{kj}$，在一定的条件下（定理 5.3.5 或定理 5.3.6），此最优解是（VOP）的非劣解. 变动参数 $\varepsilon_j$，$j=1,\cdots,p,j\neq k$，分别求解 $\mathrm{P}_k(\boldsymbol{\varepsilon}_k)$，可得一系列非劣解及相应的 Lagrange 乘子.

(3) 决策者给代替价值函数赋值　将上述每个非劣解 $\bar{x}$，目标函数值 $f_j(\bar{x})$，$j=1,\cdots,p$，及相应的 Lagrange 乘子 $\pi_{kj}(>0)$ 提供给决策者，要求他对代替价值函数 $w_{kj}$，$j=1,\cdots,p,j\neq k$，赋值.

(4) 求最终解　若对某个非劣解，决策者所赋的 $w_{kj}=0$，$j=1,\cdots,p,j\neq k$，则此非劣解是决策者满意的最终解，停止计算. 否则，用回归分析方法，建立代替价值函数的近似表达式

$$
w_{kj}=\hat{w}_{kj}(f_1,\cdots,f_{k-1},f_{k+1},\cdots,f_p),j=1,\cdots,p,j\neq k.
$$

求解方程组

$$
\hat{w}_{kj}(f_1,\cdots,f_{k-1},f_{k+1},\cdots,f_p)=0,j=1,\cdots,p,j\neq k, \tag{5.50}
$$

得到解 $(f_1^*,\cdots,f_{k-1}^*,f_{k+1}^*,\cdots,f_p^*)$，令 $\varepsilon_j=f_j^*$，$j=1,\cdots,p,j\neq k$，求解

$$
\begin{aligned} &\max\ f_k(x);\\ &\text{s.t. } f_j(x)\geqslant \varepsilon_j,j=1,\cdots,p,j\neq k,\\ &\qquad g_i(x)\leqslant 0,i=1,\cdots,m. \end{aligned}
$$

得最优解 $x^*$ 及相应的 Lagrange 乘子 $\pi_{kj}^*>0$，请决策者对 $(f_1(x^*),\cdots,f_p(x^*))$ 进行评议. 若决策者认为这是最终解，停止

计算；否则，转(2)，求更多的非劣解，重复这个过程，直到求出最终解为止.

**注 1** 若(5.50)无可行解，求更多的非劣解，继续执行这个算法.

**注 2** 若对某组参数 $\epsilon_k$，$P_k(\epsilon_k)$无解，或有最优解但出现某些$\pi_{kj}=0$，此时应抛弃这组参数，换一组参数，继续执行此算法.

**例 5.5.6** 用代替价值函数法求解

$$\max\ [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), f_3(\boldsymbol{x})];$$

$$\text{s.t. } \boldsymbol{x} \in X = \{\boldsymbol{x} = (x_1, x_2) \in E^2 \mid 0 \leqslant x_1 \leqslant 3, 0 \leqslant x_2 \leqslant 2\},$$

其中

$$f_1(\boldsymbol{x}) = -(x_1 - 3)^2 - (x_2 - 2)^2,$$

$$f_2(\boldsymbol{x}) = -x_1 - x_2,$$

$$f_3(\boldsymbol{x}) = -x_1 - 2x_2.$$

**解** (1) 选定 $f_1(\boldsymbol{x})$作为主目标函数. 求出 $f_2(\boldsymbol{x})$与 $f_3(\boldsymbol{x})$在 $X$ 上的极小值与极大值为 $m_2=-5, M_2=0, m_3=-7, M_3=0$.

(2) 求非劣解 给定 $\epsilon_2, \epsilon_3$ 的若干组值满足 $m_i \leqslant \epsilon_i \leqslant M_i, i=2, 3$. 求解约束问题

$$\max f_1(\boldsymbol{x}) = -(x_1 - 3)^2 - (x_2 - 2)^2; \tag{5.51}$$

$$\left.\begin{aligned} \text{s.t.}\quad & -x_1 - x_2 \geqslant \epsilon_2, \\ & -x_1 - 2x_2 \geqslant \epsilon_3, \\ & (x_1, x_2) \in X. \end{aligned}\right\} \tag{5.52}$$

可求非劣解. 下面给出 9 组有代表性的$(\epsilon_2, \epsilon_3)$的值及相应的最优解$(x_1, x_2)$，最优函数值 $f_1$ 及 Lagrange 乘子 $\pi_{12}$与 $\pi_{13}$(表 5.7).

(3) 给代替价值函数赋值 将上述各组$(f_1, f_2, f_3)$与相应的$(\pi_{12}, \pi_{13})$提供给决策者，决策者针对 $\pi_{12}$给 $w_{12}$赋值、针对 $\pi_{13}$给 $w_{13}$赋值. 所赋的值见表 5.7 的最后二列.

表 5.7

| $\varepsilon_2=f_2$ | $\varepsilon_3=f_3$ | $f_1$ | $x_1$ | $x_2$ | $\pi_{12}$ | $\pi_{13}$ | $w_{12}$ | $w_{13}$ |
|---|---|---|---|---|---|---|---|---|
| −4.65 | −6.45 | −0.063 | 2.85 | 1.8 | 0.2 | 0.1 | −10 | −5 |
| −4.35 | −5.975 | −0.216 | 2.725 | 1.625 | 0.35 | 0.2 | −9 | −4 |
| −4.05 | −5.5 | −0.463 | 2.6 | 1.45 | 0.5 | 0.3 | −7 | −3 |
| −3.75 | −5.025 | −0.801 | 2.475 | 1.275 | 0.65 | 0.4 | −6 | −3 |
| −3.45 | −4.55 | −1.233 | 2.35 | 1.1 | 0.8 | 0.5 | −4 | −2 |
| −3.15 | −4.075 | −1.756 | 2.225 | 0.925 | 0.95 | 0.6 | −2 | −1 |
| −2.85 | −3.6 | −2.373 | 2.1 | 0.75 | 1.1 | 0.7 | −1 | 0 |
| −2.55 | −3.125 | −3.081 | 1.975 | 0.575 | 1.25 | 0.8 | 1 | 0 |
| −2.25 | −2.65 | −3.883 | 1.85 | 0.4 | 1.4 | 0.9 | 3 | 1 |

(4) 求最终解　因为表中各非劣解的 $w_{12}$ 与 $w_{13}$ 不同时等于 0，这些非劣解不是最终解. 根据表 5.7 的数据，可以用多元线性回归构造出函数 $w_{12}(f_2,f_3)$ 与 $w_{13}(f_2,f_3)$，它们是

$$w_{12}(f_2,f_3)=13.344+1.4f_2+2.744f_3,$$

$$w_{13}(f_2,f_3)=6.641+1.75f_2+0.5442f_3.$$

令 $w_{12}(f_2,f_3)=0$ 与 $w_{13}(f_2,f_3)=0$，解出

$$f_2^*=-2.72, f_3^*=-3.47.$$

用 $\varepsilon_2=f_2^*=-2.72$，$\varepsilon_3=f_3^*=-3.47$ 代入约束问题(5.51)、(5.52)中的 $\varepsilon_2$ 与 $\varepsilon_3$，并求解这个问题，得到 $\boldsymbol{x}^*=(1.97,0.75)$，相应的 $(f_1^*,f_2^*,f_3^*)=(-2.623,-2.72,-1.47)$，决策者表示满意，$\boldsymbol{x}^*$ 是最终解.

## 5.6 目标规划

一种特殊的多目标规划叫**目标规划**(goal programming)，这是美国学者 Charnes 等在 1952 年提出来的. 目标规划的重要特点是对各个目标分级加权与逐级优化，这符合人们处理问题要分别轻重缓

急、保证重点的思考方式，是目前应用很广的一种多目标规划.

### 5.6.1 目标规划的一般形式

目标规划与一般多目标规划在形式上有差别，先举例说明.

**例 5.6.1** 某木工小组生产椅子与桌子两种产品，每周生产时间为 48 小时，生产一把椅子平均需要 0.4 小时，生产一张桌子需要 1 小时. 根据市场预测，椅子的销售量为每周 60 把，桌子的销售量为每周 30 张. 每把椅子的利润为 8 元，每张桌子的利润为 20 元. 在制定生产计划时，组长提出下述 4 项目标：

(1) 总利润最大；

(2) 工人加班时间最少；

(3) 产量不能超过市场预测的销售量；

(4) 尽可能满足市场需求.

试建立这个问题的数学模型.

**解** 设决策变量为

$x_1$ = 每周椅子的产量；

$x_2$ = 每周桌子的产量.

在建立模型时，与组长(决策者)讨论这 4 个目标的优先次序是：产量不超过市场预测，工人加班时少，总利润最大，满足市场需求.

第一优先级目标为产量不超过市场预测，即 $x_1 \leqslant 60$，$x_2 \leqslant 30$. 我们引进偏差变量 $d_1^+$，$d_1^-$，$d_2^+$，$d_2^-$，把上述两个不等式改写为等式约束条件

$$x_1 + d_1^- - d_1^+ = 60, x_2 + d_2^- - d_1^+ = 30,$$

其中 $d_1^+$ 与 $d_2^+$ 叫**正偏差变量**(positive deviational variables)，$d_1^-$ 与 $d_2^-$ 叫**负偏差变量**(negative deviational variables)，它们的含意是

$d_1^+$ = 椅子产量 $x_1$ 超过 60 把的偏差量；

$d_1^-$ = 椅子产量 $x_1$ 不足 60 把的偏差量；

$d_2^+$＝桌子产量 $x_2$ 超过 30 张的偏差量；

$d_2^-$＝桌子产量 $x_2$ 不足 30 张的偏差量.

显然，$d_i^+\geqslant 0$，$d_i^-\geqslant 0$，且 $d_i^+d_i^-=0$，$i=1,2$，即 $d_i^+$ 与 $d_i^-$ 中必有一个等于 0. 例如，若椅子产量为 $x_1=70$ 时，则 $d_1^+=10$，$d_1^-=0$；若 $x_1=55$，则 $d_1^+=0$，$d_1^-=5$. 引进偏差变量以后，可将目标 $x_1\leqslant 60$ 与 $x_2\leqslant 30$ 表示成两个极小化问题 $\min d_1^+$ 与 $\min d_2^+$，并且可以合并为

$$\min\ (d_1^+ + d_2^+).$$

这是第一优先级需要完成的任务.

第二优先级的目标是工人加班时间最少. 因椅子与桌子的每周产量是 $x_1$ 与 $x_2$，工人每周工作时间为 $0.4x_1+x_2$ 小时，但规定的标准是每周工作 48 小时，即

$$0.4x_1+x_2+d_3^- - d_3^+ = 48,$$

其中 $d_3^+$ 为工作超过 48 小时的偏差量，$d_3^-$ 为工作不足 48 小时的偏差量. 决策者希望工人加班的时间最少，即第二优先级需要完成的任务是

$$\min d_3^+.$$

第三优先级的目标是总利润最大. 当产量是 $x_1$ 与 $x_2$ 时，总利润为 $8x_1+20x_2$，假设决策者想像中的最高利润为 1500 元，建立约束条件

$$8x_1+20x_2+d_4^- - d_4^+ = 1500.$$

决策者希望 $d_4^-$ 尽可能小，即

$$\min d_4^-,$$

这是第三优先级需要完成的任务.

第四优先级目标是希望产量 $x_1$ 与 $x_2$ 尽可能满足市场需求，从前面已经建立的两个约束条件

$$x_1+d_1^- - d_1^+ = 60,\quad x_2+d_2^- - d_2^+ = 30,$$

可以看出，这要求 $d_1^-$ 与 $d_2^-$ 尽可能小. 另一方面从顾客的角度看，每张桌子的重要性是每把椅子重要性的两倍，因此希望减少 $d_2^-$ 的愿望也是减少 $d_1^-$ 愿望的两倍. 统一考虑，第四优先级需要完成的任务是

$$\min(d_1^- + 2d_2^-).$$

这里，称系数 2 是 $d_2^-$ 的权因子，$d_1^-$ 的权因子是 1.

将上述四个极小问题按优先级顺序统一写成

$$\min\{a_1, a_2, a_3, a_4\} = \{(d_1^+ + d_2^+), d_3^+, d_4^-, (d_1^- + 2d_2^-)\}.$$

总结起来，这个问题的目标规划模型是：

$$\min\{a_1, a_2, a_3, a_4\} = \{(d_1^+ + d_2^+), d_3^+, d_4^-, (d_1^- + 2d_2^-)\}; \tag{5.53}$$

$$\left.\begin{aligned} \text{s.t.} \quad x_1 \qquad\quad + d_1^- - d_1^+ &= 60, \\ x_2 + d_2^- - d_2^+ &= 30, \\ 0.4x_1 + x_2 + d_3^- - d_3^+ &= 48, \\ 8x_1 + 20x_2 + d_4^- - d_4^+ &= 1500, \\ x_1, x_2 &\geqslant 0, \\ d_i^+, d_i^- &\geqslant 0, i = 1,2,3,4. \end{aligned}\right\} \tag{5.54}$$

目标规划的一般形式是

$$\min(a_1, \cdots, a_K) = \left\{\sum_{i=1}^m v_{1i} d_i^- + u_{1i} d_i^+, \cdots, \sum_{i=1}^m v_{Ki} d_i^- + u_{Ki} d_i^+\right\}; \tag{5.55}$$

$$\left.\begin{aligned} \text{s.t.} \quad f_i(\boldsymbol{x}) + d_i^- - d_i^+ = b_i, i &= 1, \cdots, m, \\ x_j \geqslant 0, j &= 1, \cdots, n, \\ d_i^-, d_i^+ \geqslant 0, i &= 1, \cdots, m. \end{aligned}\right\} \tag{5.56}$$

这里 $x_j, j=1,\cdots,n$，称为**决策变量**(decision variables)，$d_i^+$ 与 $d_i^-$，$i=1,\cdots,m$，分别是正偏差变量与负偏差变量，$a_k = \sum_{i=1}^m v_{ki} d_i^- +$

$u_{ki}d_i^+$, $k=1,\cdots,K$，是 $d_i^-$ 与 $d_i^+$ 的线性函数，称为**完成函数**(achievement functions)，其中 $k$ 是优先级，$k$ 的值愈小，表示 $a_k$ 愈重要. 极小化要逐级进行，在优化 $k$ 级时，不允许破坏第 1，…，第 $k-1$ 级的最优性，$K$ 是优先级的个数，系数 $v_{ki}$ 与 $u_{ki}$ 称为权因子. $f_i(\boldsymbol{x})$ 称为目标函数，$b_i$ 称为目标值，规定 $b_i\geqslant 0$. 称 $f_i(\boldsymbol{x})+d_i^- - d_i^+ = b_i$ 为约束条件.

当所有的目标函数都是线性函数时，即 $f_i(\boldsymbol{x}) = \sum_{j=1}^{n} a_{ij}x_j$, $i=1,\cdots,m$ 时，称这种目标规划为**线性目标规划**(linear goal programming)；若 $f_i(\boldsymbol{x})$ 中有非线性函数时，则称为**非线性目标规划**.

在考虑用 $f_i(\boldsymbol{x})+d_i^- - d_i^+ = b_i$ 来建立完成函数时，会有三种可能的愿望(1) $f_i(\boldsymbol{x})\geqslant b_i$，或(2) $f_i(\boldsymbol{x})\leqslant b_i$，或(3) $f_i(\boldsymbol{x})=b_i$. 这三种可能性分别可以用极小化偏差变量的一个线性函数(见表 5.8)来完成.

表 5.8

| | |
|---|---|
| $f_i(\boldsymbol{x})\geqslant b_i$ | $\min d_i^-$ |
| $f_i(\boldsymbol{x})\leqslant b_i$ | $\min d_i^+$ |
| $f_i(\boldsymbol{x})=b_i$ | $\min(d_i^+ + d_i^-)$ |

要注意，只有当 $\min d_i^- = 0$ 时，才能实现 $f_i(\boldsymbol{x})\geqslant b_i$；否则若 $\min d_i^- > 0$，就有 $f_i(\boldsymbol{x})<b_i$，决策者的愿望不能实现. 其他情况也类似.

### 5.6.2 目标规划的图解法

可以用图解法求解两个决策变量的线性目标规划问题.

图解法的计算步骤：

(1) 设有两个决策变量的线性目标规划为

$$\min\{a_1,\cdots,a_K\}=\left\{\sum_{i=1}^{m}v_{1i}d_i^- + u_{1i}d_i^+,\cdots,\sum_{i=1}^{m}v_{Ki}d_i^- + u_{Ki}d_i^+\right\};$$

$$\text{s.t.}\quad a_{i1}x_1 + a_{i2}x_2 + d_i^- - d_i^+ = b_i, i=1,\cdots,m,$$

$$x_1,x_2\geqslant 0,$$

$$d_i^-,d_i^+\geqslant 0, i=1,\cdots,m.$$

在平面上画出直线 $a_{i1}x_1+a_{i2}x_2=b_i, i=1,\cdots,m$,的图形.

(2) 求出第一优先级完成函数 $a_1$ 的最优解集合.

(3) 按顺序求出 $a_2,a_3,\cdots,a_K$ 各级的最优解集合,求第 $k$ 级最优解时,不能破坏以前各级($1,2,\cdots,k-1$ 级)的最优性. 第 $K$ 级的最优解就是目标规划的最优解.

**例 5.6.2** 用图解法求解下述线性目标规划

$$\min\{a_1,a_2,a_3,a_4\}=\{(d_1^+ + d_2^+),d_3^+,d_4^-,(d_1^- + 1.5d_2^-)\};$$

$$\begin{aligned}\text{s.t.}\ x_1 \qquad\quad + d_1^- - d_1^+ &= 20,\\ x_2 + d_2^- - d_2^+ &= 40,\\ x_1 + x_2 + d_3^- - d_3^+ &= 50,\\ x_1 + x_2 + d_4^- - d_4^+ &= 80,\\ x_1,x_2 &\geqslant 0,\\ d_j^+,d_j^- \geqslant 0, j &= 1,2,3,4.\end{aligned}$$

**解** 求解过程用图 5.17 到图 5.20 来表示,图 5.17 中的 4 条直线分别代表 4 个约束条件,在图 5.18 中考虑第一优先级 $\min(d_1^+ + d_2^+)$,令 $d_1^+=0, d_2^+=0$,画阴影部分满足这个要求,是这个极小值问题的最优解集合.

然后,在第一优先级的最优解集合中,求第二优先级的 $\min d_3^+$ 的最优解,在图 5.18 画阴影部分中找出满足 $d_3^+=0$ 的点,就得到图 5.19 中画阴影部分,这是第二优先级的最优解集合.

考虑第三优先级 $\min d_4^-$ 时,在不破坏第一、第二优先级的最优性前提下使 $d_4^-$ 最小,从图 5.19 的画阴影部分找出使 $d_4^-$ 最小

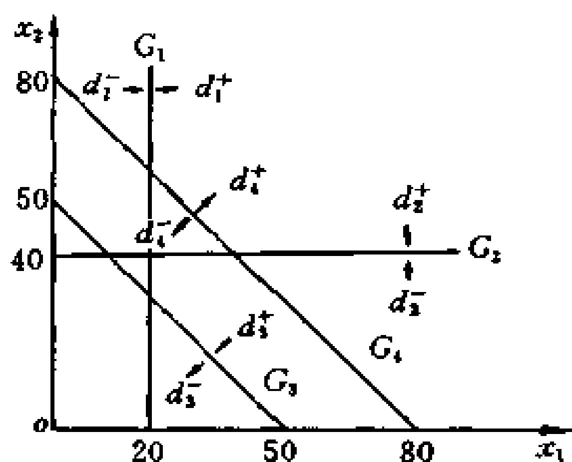



图 5.17

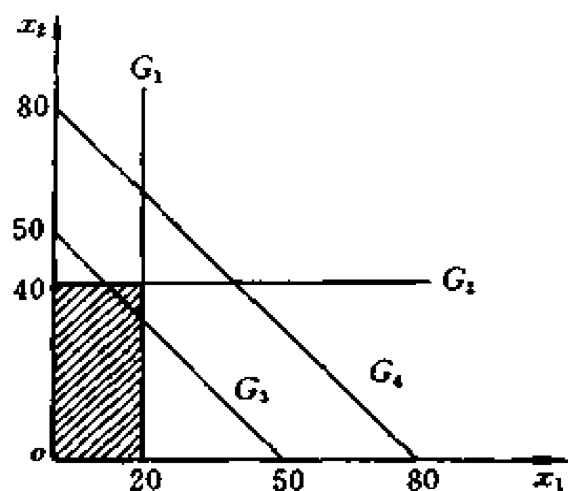



图 5.18

的点. 因为画阴影部分与直线 $G_4$ 不相交,不可能使 $d_4^-=0$. 线段 $AB$ 上所有的点是使 $d_4^-$ 取极小值的点. 图 5.20 中的 $AB$ 线段表示第三优先级最优解集合.

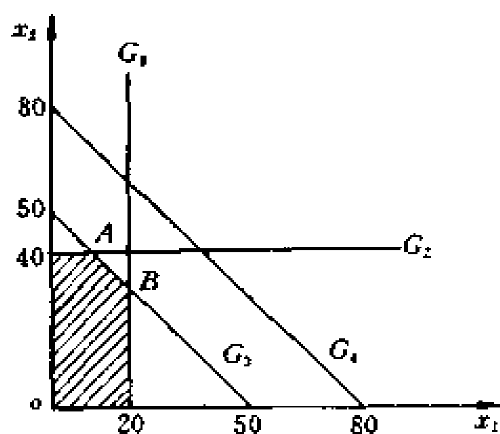



图 5.19

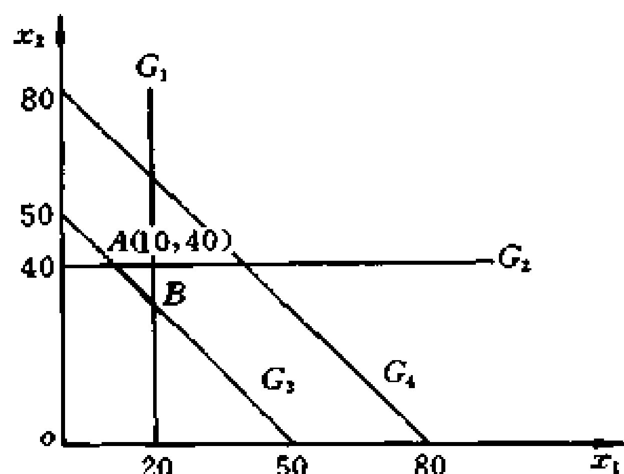



图 5.20

最后,考虑第四优先级 $\min(d_1^-+1.5d_2^-)$,要在线段 $AB$ 上找出 $\min(d_1^-+1.5d_2^-)$的解. 希望 $d_1^-$ 与 $d_2^-$ 都尽可能小,但 $d_2^-$ 的权因子 1.5 比 $d_1^-$ 的权因子 1 大,$d_2^-$ 比 $d_1^-$ 重要,令 $d_2^-=0$,得到 $A$

点是最优解(图 5.20). $A$ 点的坐标是(10,40),此目标规划的最优解是 $x_1^*=10, x_2^*=40$.

此时各级完成函数值是

$$(a_1,a_2,a_3,a_4)=(d_1^+ + d_2^+, d_3^+, d_4^-, d_1^- + 1.5d_2^-)$$
$$=(0,0,30,10).$$

这表明第一、第二级完成函数值都等于零,第三、第四级完成函数值大于零. 因此,决策者第一、二两项愿望能够实现. 第三、四两项愿望不能完全实现,但$(x_1^*, x_2^*)$已经是最好的结果了.

### 5.6.3 目标规划的单纯形法

从上述图解法可以看出,求解线性目标规划实际上是求解多级线性规划. 若每级线性规划都用单纯形法求解就形成目标规划的单纯形法.

设线性目标规划为

$$\text{(GLP)}\quad \min\{a_1,\cdots,a_K\} = \left\{\sum_{i=1}^{m} v_{1i}d_i^- + u_{1i}d_i^+, \cdots, \sum_{i=1}^{m} v_{Ki}d_i^- + u_{Ki}d_i^+\right\}; \tag{5.57}$$

$$\left.\begin{aligned} \text{s.t.}\ & \sum_{j=1}^{n} a_{ij}x_j + d_i^- - d_i^+ = b_i, i = 1,\cdots,m, \\ & x_j \geqslant 0, d_i^+, d_i^- \geqslant 0, j = 1,\cdots,n, i = 1,\cdots,m \end{aligned}\right\} \tag{5.58}$$

用单纯形法求解,首先要建立单纯形表,此表与线性规划的单纯形表相似,不同的是,因有 $K$ 级完成函数,表中有 $K$ 行写完成函数的系数 $v_{ki}$与 $u_{ki}$(相当于线性规划的目标函数的系数),及与这 $K$ 级完成函数对应的 $K$ 行检验数. 设初始单纯形表是表 5.9.

**表 5.9　初始单纯形表**

| | | | $P_1$ | 0 | ⋯ | 0 | $v_{11}$ | ⋯ | $v_{1m}$ | $u_{11}$ | ⋯ | $u_{1m}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | |
| | | | $P_K$ | 0 | ⋯ | 0 | $v_{K1}$ | ⋯ | $v_{Km}$ | $u_{K1}$ | ⋯ | $u_{Km}$ | |
| $P_1$ | ⋯ | $P_K$ | | $x_1$ | | $x_n$ | $d_1^-$ | ⋯ | $d_m^-$ | $d_1^+$ | ⋯ | $d_m^+$ | $b$ |
| $v_{11}$ | ⋯ | $v_{K1}$ | $d_1^-$ | $a_{11}$ | ⋯ | $a_{1n}$ | $a_{1,n+1}$ | ⋯ | $a_{1,n+m}$ | $a_{1,n+m+1}$ | ⋯ | $a_{1,n+2m}$ | $b_1$ |
| ⋮ | | ⋮ | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | ⋮ |
| $v_{1m}$ | ⋯ | $v_{Km}$ | $d_m^-$ | $a_{m1}$ | ⋯ | $a_{mn}$ | $a_{m,n+1}$ | ⋯ | $a_{m,n+m}$ | $a_{m,n+m+1}$ | ⋯ | $a_{m,n+2m}$ | $b_m$ |
| | | | $P_1$ | $l_{11}$ | ⋯ | $l_{1n}$ | $l_{1,n+1}$ | ⋯ | $l_{1,n+m}$ | $l_{1,n+m+1}$ | ⋯ | $l_{1,n+2m}$ | $a_1$ |
| | | | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | ⋮ | | ⋮ | ⋮ |
| | | | $P_K$ | $l_{K1}$ | ⋯ | $l_{Kn}$ | $l_{K,n+1}$ | ⋯ | $l_{K,n+m}$ | $l_{K,n+m+1}$ | ⋯ | $l_{K,n+2m}$ | $a_K$ |

其中 $P$ 是 Priority（优先级）的简写，$P_k$ 代表第 $k$ 优先级. 初始单纯形表中，基矩阵是

$$\begin{bmatrix} a_{1,n+1} & \cdots & a_{1,n+m} \\ \vdots & & \vdots \\ a_{m,n+1} & \cdots & a_{m,n+m} \end{bmatrix} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix},$$

$d_1^-,\cdots,d_m^-$ 是基变量，其值是 $d_j^- = b_j, j=1,\cdots,m$.

$$\begin{bmatrix} a_{1,n+m+1} & \cdots & a_{1,n+2m} \\ \vdots & & \vdots \\ a_{m,n+m+1} & \cdots & a_{m,n+2m} \end{bmatrix} = \begin{bmatrix} -1 & & \\ & \ddots & \\ & & -1 \end{bmatrix}.$$

表的上端是各级完成函数的权因子（系数）$v_{ki}$ 与 $u_{ki}$，同时，在表的左端也列出基变量 $d_i^-$ 的系数，表的下部是对应各级完成函数的检验数 $l_{kj}$ 与相应的完成函数 $a_k$ 的值，$l_{kj}$ 的计算公式与线性规划的相同，为

$$l_{kj} = \begin{cases} \sum_{i=1}^{m} v_{ki} a_{ij}, & \text{当 } j = 1,\cdots,n, \\ 0 & \text{当 } j = n+1,\cdots,n+m, \\ \sum_{i=1}^{m} v_{ki} a_{ij} - u_{k,j-n-m}, & \text{当 } j = n+m+1,\cdots,n+2m, \end{cases}$$

$$k=1,\cdots,K.$$

对应于基变量的检验数都等于 0,即

$$\begin{bmatrix} l_{1,n+1} & \cdots & l_{1,n+m} \\ \vdots & & \vdots \\ l_{K,n+1} & \cdots & l_{K,n+m} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}.$$

**例 5.6.3** 列出下面目标规划的初始单纯形表

$$\min\{d_1^- + d_2^+, d_3^-, 2d_2^- + d_3^+\}; \tag{5.59}$$

$$\left.\begin{aligned} \text{s.t.}\quad x_1 \quad + d_1^- \quad - d_1^+ &= 10, \\ 2x_1 + x_2 \quad + d_2^- \quad - d_2^+ &= 40, \\ 3x_1 + 2x_2 \quad + d_3^- \quad - d_3^+ &= 100, \\ x_1, x_2 &\geqslant 0, \\ d_1^-, d_2^-, d_3^-, d_1^+, d_2^+, d_3^+ &\geqslant 0. \end{aligned}\right\} \tag{5.60}$$

**解** 初始单纯形表为

**表 5.10**

| | | | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $P_1$ | | | 1 | | | | 1 | | |
| | | | $P_2$ | | | | | 1 | | | | |
| | | | $P_3$ | | | | 2 | | | | 1 | |
| $P_1$ | $P_2$ | $P_3$ | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $\bar{b}$ |
| 1 | | | $d_1^-$ | 1 | | 1 | | | −1 | | | 10 |
| | | 2 | $d_2^-$ | 2 | 1 | | 1 | | | −1 | | 40 |
| | 1 | | $d_3^-$ | 3 | 2 | | | 1 | | | −1 | 100 |
| | | | $P_1$ | 1 | | | | | −1 | −1 | | 10 |
| | | | $P_2$ | 3 | 2 | | | | | | −1 | 100 |
| | | | $P_3$ | 4 | 2 | | | | | −2 | −1 | 80 |

这里,基变量是 $d_1^-, d_2^-$ 与 $d_3^-$,其值为 $d_1^-=10, d_2^-=40, d_3^-=100$,其它自变量都是非基变量,都取 0 值,完成函数的值是 $a_1=10, a_2=100, a_3=80$.

单纯形法的计算步骤：

(1) 建立初始单纯形表，令 $k=1$.

(2) 检查第 $k$ 级完成函数的最优性　若 $a_k=0$，则第 $k$ 级完成函数已达到最优，转(6)，否则，检查第 $k$ 行检验数 $l_{ks}$，找出满足

$$\begin{cases} \text{条件 1} \quad l_{ks} > 0, \\ \text{条件 2} \quad l_{1s}, \cdots, l_{k-1,s} \text{ 中没有负数} \end{cases}$$

的所有 $l_{ks}$，再求

$$l_{ks'} = \max_s \{ l_{ks} | l_{ks} > 0, l_{1s}, \cdots, l_{k-1,s} \text{ 中没有负数} \}.$$

若有几个 $l_{ks'}$ 有相同的值，任选其中一个，转(3)；若不存在满足条件 1,2 的 $l_{ks}$，表示第 $k$ 级完成函数已达到最优，转(6).

(3) 令第 $s'$ 列的变量为进基变量.

(4) 确定出基变量，求出

$$\frac{b_{i'}}{a_{i's'}} = \min_i \left\{ \frac{b_i}{a_{is'}} \middle| a_{is'} > 0 \right\}.$$

令第 $i'$ 行的变量为出基变量. 若 $i$ 不止一个，任选其中一个(在理论上说，可能导致循环，但实际上几乎不可能发生).

(5) 建立新的单纯形表　用消去变换将第 $s'$ 列变为单位向量. 具体做法是：

1) 将第 $i'$ 行的所有元素除以 $a_{i's'}$ 后，放入新表的第 $i'$ 行，即

$$\bar{a}_{i's} = \frac{a_{i's}}{a_{i's'}}, \ s = 1, \cdots, n + 2m,$$

$$\bar{b}_{i'} = \frac{b_{i'}}{a_{i's'}}.$$

2) 将所有各行(包括检验数行，但第 $i'$ 行除外)加上 $i'$ 行的若干倍，使此行第 $s'$ 列处的元素变为 0，将结果放入新表，即令

$$\begin{cases} \bar{a}_{is} = a_{is} - \dfrac{a_{is'}}{a_{i's'}} a_{i's}, & i = 1, \cdots, m, i \neq i', \\ \bar{b}_i = b_i - \dfrac{a_{is'}}{a_{i's'}} b'_i, & s = 1, \cdots, n + 2m, \end{cases}$$

$$\begin{cases} \bar{l}_{ks} = l_{ks} - \dfrac{l_{ks'}}{a_{i's'}} a_{i's}, & k = 1,\cdots,K, \\ \bar{a}_k = a_k - \dfrac{l_{ks'}}{a_{i's'}} a_{i'}, & s = 1,\cdots,n+2m. \end{cases}$$

转(2).

(6) 准备进行下一优先级的优化　令 $k=k+1$,若 $k>K$,则停止计算,已经得到最优解;否则,转(2).

**例 5.6.4**　用单纯形法解线性目标规划(5.53)、(5.54)

$$\min\{a_1,a_2,a_3,a_4\} = \{(d_1^+ + d_2^+), d_3^+, d_4^-, (d_1^- + 2d_2^-)\};$$

$$\begin{aligned} \text{s.t.}\quad x_1 \qquad\quad + d_1^- - d_1^+ &= 60, \\ x_2 + d_2^- - d_2^+ &= 30, \\ 0.4x_1 + x_2 + d_3^- - d_3^+ &= 48, \\ 8x_1 + 20x_2 + d_4^- - d_4^+ &= 1500, \\ x_1, x_2 &\geqslant 0, \\ d_i^+, d_i^- \geqslant 0, i &= 1,\cdots,4. \end{aligned}$$

**解**　此问题有 2 个决策变量与 8 个偏差变量,$n=2$,$m=4$,完成函数有 4 个优先级,$K=4$.

(1) 建立初始单纯形表 5.11. 这里,基变量为 $d_i^-$,$i=1,\cdots,4$,$(d_1^-,d_2^-,d_3^-,d_4^-)=(60,30,48,1500)$,完成函数值$(a_1,a_2,a_3,a_4)=(0,0,1500,120)$. 令 $k=1$.

(2) 因为 $a_1=0$,第一优先级已达到最优,转(6).

(6) $k=k+1=2$,因为 $k<K(=4)$,转(2).

(2) 因为 $a_2=0$,第二优先级也达到最优,转(6).

(6) $k=k+1=3$,因为 $k<K$,转(2).

(2) 因 $a_3=1500>0$,检查第 $k(=3)$行检验数,此行有 2 个检验数 8 与 20 符合条件 1,2. 求出其中的极大值

$$l_{ks'} = \max\{8,20\} = 20,$$

这里 $k=3$,$s'=2$,转(3).

**表 5.11　初始单纯形表**

| | | | | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_4^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $d_4^+$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_1$ | | | | | | | 1 | 1 | | | |
| | | | | $P_2$ | | | | | | | | | 1 | | |
| | | | | $P_3$ | | | | | | 1 | | | | | |
| | | | | $P_4$ | | | 1 | 2 | | | | | | | |
| $P_1$ | $P_2$ | $P_3$ | $P_4$ | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_4^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $d_4^+$ | $\bar{b}$ |
| | | | 1 | $d_1^-$ | 1 | | 1 | | | | $-1$ | | | | 60 |
| | | | 2 | $d_2^-$ | | ① | | 1 | | | | $-1$ | | | 30 |
| | | | | $d_3^-$ | 0.4 | 1 | | | 1 | | | | $-1$ | | 48 |
| | | 1 | | $d_4^-$ | 8 | 20 | | | | 1 | | | | $-1$ | 1500 |
| | | | | $P_1$ | | | | | | | $-1$ | $-1$ | | | 0 |
| | | | | $P_2$ | | | | | | | | | $-1$ | | 0 |
| | | | | $P_3$ | 8 | 20 | | | | | | | | $-1$ | 1500 |
| | | | | $P_4$ | 1 | 2 | | | | | $-1$ | $-2$ | | | 120 |

(3) 令第 $s'(=2)$列非基变量 $x_2$ 进基.

(4) 求出

$$\frac{b_{i'}}{a_{i's'}}=\min\left\{\frac{30}{1},\frac{48}{1},\frac{1500}{20}\right\}=\frac{30}{1},$$

这里 $i'=2,s'=2$,令第 $i'(=2)$行变量 $d_2^-$ 出基.

(5) 做消去变换,建立新单纯形表.

1) 因为 $a_{i's'}=a_{22}=1$,第 2 行的元素不动,基变量改为 $x_2$,因各级完成函数中 $x_2$ 的权因子都是 0,新表的右端也做相应的改动.

2) 用第 2 行乘以$-1$,加到第 3 行;用第 2 行乘以$-20$ 加到第 4 行;用第 2 行乘以$-20$ 加到 $P_3$ 行(即第 3 行检验数);用第 2 行乘以$-2$ 加到 $P_4$ 行(即第 4 行检验数);其它行不变,得到新单纯形表 5.12,转(2).

(2) 检查表 5.12 的第 $k(=3)$行的检验数,符合条件 1,2 的只有 $l_{ks'}=8$,这里 $k=3,s'=1$.

表 5.12

| $P_1$ | $P_2$ | $P_3$ | $P_4$ | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_4^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $d_4^+$ | $b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_1$ | | | | | | | 1 | 1 | | | |
| | | | | $P_2$ | | | | | | | | | 1 | | |
| | | | | $P_3$ | | | | | | 1 | | | | | |
| | | | | $P_4$ | | | 1 | 2 | | | | | | | |
| | | | 1 | $d_1^-$ | 1 | | 1 | | | | −1 | | | | 60 |
| | | | | $x_2$ | | 1 | | 1 | | | | −1 | | | 30 |
| | | | | $d_3^-$ | (0.4) | | | −1 | 1 | | | 1 | −1 | | 18 |
| | | 1 | | $d_4^-$ | 8 | | | −20 | | 1 | | 20 | | −1 | 900 |
| | | | | $P_1$ | | | | | | | −1 | −1 | | | 0 |
| | | | | $P_2$ | | | | | | | | | −1 | | 0 |
| | | | | $P_3$ | 8 | | | −20 | | | | 20 | | −1 | 900 |
| | | | | $P_4$ | 1 | | | −2 | | | −1 | | | | 60 |

(3) 令第 $s'(=1)$列的非基变量 $x_1$ 进基.

(4) 确定出基变量.计算

$$\frac{b_{i'}}{a_{i's'}} = \min\left\{\frac{60}{1},\frac{18}{0.4},\frac{900}{8}\right\} = \frac{18}{0.4},$$

这里 $i'=3,s'=1$.令第 3 行的基变量 $d_3^-$ 出基.

(5) 做消去变换,建立新单纯形表.以 0.4 为主元素,对单纯形表 5.12 做消去变换,得到新单纯形表 5.13,转(2).

(2) 检查表 5.13 的 $P_3$ 行检验数,这一行的检验数都不符合条件 1,2.第 3 级达到最优,转(6).

(6) $k=k+1=4,k\leqslant K(=4)$,转(2).

(2) 检查表 5.13 的 $P_4$ 行检验数,只有 $l_{ks'}=0.5$ 符合条件 1,2,这里 $k=4,s'=4$.

(3) 令第 $s'(=4)$列的非基变量 $d_2^-$ 进基.

(4) 计算

$$\frac{b_{i'}}{a_{i's'}} = \min\left\{\frac{15}{2.5},\frac{30}{1}\right\} = \frac{15}{2.5},$$

表 5.13

| $P_1$ | $P_2$ | $P_3$ | $P_4$ | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_4^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $d_4^+$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_1$ | | | | | | | 1 | 1 | | | |
| | | | | $P_2$ | | | | | | | | | 1 | | |
| | | | | $P_3$ | | | | | | 1 | | | | | |
| | | | | $P_4$ | | | 1 | 2 | | | | | | | |
| | | | 1 | $d_1^-$ | | | 1 | (2.5) | −2.5 | | −1 | −2.5 | 2.5 | | 15 |
| | | | | $x_2$ | | 1 | | 1 | | | | −1 | | | 30 |
| | | | | $x_1$ | 1 | | | −2.5 | 2.5 | | | 2.5 | −2.5 | | 45 |
| | | 1 | | $d_4^-$ | | | | | −20 | 1 | | | 20 | −1 | 540 |
| | | | | $P_1$ | | | | | | | −1 | −1 | | | 0 |
| | | | | $P_2$ | | | | | | | | | −1 | | 0 |
| | | | | $P_3$ | | | | | −20 | | | | 20 | −1 | 540 |
| | | | | $P_4$ | | | | 0.5 | −2.5 | | −1 | −2.5 | 2.5 | | 15 |

这里 $i'=1, s'=4$. 令第 1 行的基变量 $d_1^-$ 出基.

(5) 做消去变换、建立新单纯形表. 以第 1 行第 4 列元素 2.5 作主元素，对表 5.13 做消去变换，得到新的单纯形表 5.14，转(2).

表 5.14

| $P_1$ | $P_2$ | $P_3$ | $P_4$ | | $x_1$ | $x_2$ | $d_1^-$ | $d_2^-$ | $d_3^-$ | $d_4^-$ | $d_1^+$ | $d_2^+$ | $d_3^+$ | $d_4^+$ | $\bar{b}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $P_1$ | | | | | | | 1 | 1 | | | |
| | | | | $P_2$ | | | | | | | | | 1 | | |
| | | | | $P_3$ | | | | | | 1 | | | | | |
| | | | | $P_4$ | | | 1 | 2 | | | | | | | |
| | | | 2 | $d_2^-$ | | | 0.4 | 1 | −1 | | −0.4 | −1 | 1 | | 6 |
| | | | | $x_2$ | | 1 | −0.4 | | 1 | | 0.4 | | −1 | | 24 |
| | | | | $x_1$ | 1 | | 1 | | | | −1 | | | | 60 |
| | | 1 | | $d_4^-$ | | | | | −20 | 1 | | | 20 | −1 | 540 |
| | | | | $P_1$ | | | | | | | −1 | −1 | | | 0 |
| | | | | $P_2$ | | | | | | | | | −1 | | 0 |
| | | | | $P_3$ | | | | | −20 | | | | 20 | −1 | 540 |
| | | | | $P_4$ | | | −0.2 | | −2 | | −0.8 | −2 | 2 | | 12 |

（2）检查表 5.14 中 $P_4$ 行的检验数．这一行的检验数都不满足条件 1，2．第 4 级已经达到最优，转（6）．

（6）$k=k+1=5$，因为 $k>K(=4)$，已经找到最优解，停止计算．

最后找到的最优解是 $x_1^*=60$，$x_2^*=24$．在例 5.6.1 中，此最优解表示，这个木工小组每周应该生产椅子 60 把、桌子 24 张．在单纯形表 5.14 中，各级完成函数值是

$$(a_1,a_2,a_3,a_4)=(d_1^+ + d_2^+,d_3^+,d_4^-,d_1^- + 2d_2^-)$$
$$=(0,0,540,12).$$

这表明，木工小组的第一优先级的目标，即产量不超过市场预测量，可以完成．第二优先级目标，即工人加班时间尽量少，也可以完成．第三优先级的目标，即总利润为 1500 元，没有达到，还差 540 元．第四优先级的目标，即尽量满足市场需求，没有达到．虽然如此，但在规定的优先级顺序下，这组解已经是最好的了．

## 参考资料


1 魏权龄等．数学规划与优化设计．国防工业出版社，1984

2 Benayoun R, et al. Linear Programming with Multiple Objective Functions. STEP Method (STEM) Math. Programming 1, 366～375, 1971

3 Changkong V, Haimes Y Y. Multiobjective Decision Making: Theory and Methodology. North-Holland Publ., 1983

4 Cohon J L. Multiobjective Programming and Planning. Academic Press, New York, 1978

5 Huang C L., Masud A S M. Multiobjective Decision Making Methods and Applications. Springer-Verlag, 1979

6 Ignizio J P. Goal Programming and Extensions. Lexington Books, 1976

7 Lee S M. Goal Programming for Decision Analysis. Auerbah Publ., 1972

8 Steuer R E. Multiple Criteria Optimization: Theory, Computation, and Applications. John Wiley & Sons, 1986

9 Zionts S, Wallenius J. An Iterative Method for Solving the Multiple Criteria Problem. Management Science, 22.6, 652~663, 1976

# 6 对策论

## 6.1 矩阵对策

### 6.1.1 引言

对策论的发展和研究内容

**对策论**(game theory) 成为数学的一个分支，始于 1944 年. 这一年，J. von Neumann 和 O. Morgenstern 出版了奠基性的经典著作 Theory of Games and Economic Behavior，第一次给**对策**(game)以明确的数学描述，对有关对策的一些理论作出了系统的论证，并且讨论了对策在经济学上的一些应用.

对策论是研究两个或两个以上参加者在对抗性或竞争性局势下如何采取行动，作出有利于己方的决策的数学理论. 也就是在两个或多个参加者相互之间利害有冲突的情况下，各个参加者应如何分析各方面的局势，权衡利弊，以决定采取怎样的行动的数学理论. 各方面的利害关系可为对抗性或竞争性的，每方只能控制部分局势(变量)，有时也不排除部分参加者之间有所合作，最终的目的是各自作出抉择，以期得到一个对己方最有利的结局.

从 1944 年到现在，对策论在理论和应用方面都有了不少发展. 在理论方面，从最初的**零和二人对策**(zero-sum two-person game)发展到**非零和 $n$ 人对策**(non-zero-sum n-person game)；最近 10 多年来特别在 **$n$ 人合作对策**(n-person cooperative game)方面的研究有很大的进展. 在应用方面，从最初的经济学领域扩展到军事、政治、社会学、心理学等问题，近年来又回到了经济学方面；

但主要还是定性的研究.

**矩阵对策**(matrix game)是最基本的一类对策. 为了有比较直观的了解,首先通过几个简单的例子来说明一些有关的基本概念.

**例 6.1.1** 配钱币游戏

两个参加者,称为**局中人**(player) 1 和 2,各拿出一枚钱币. 在不让对方看见的情况下,将钱币出示给对方. 如果两个钱币都呈正面或都呈反面,则局中人 1 得 1 分,局中人 2 得 $-1$ 分;或者说,局中人 2 输给局中人 1 一个单位. 如果两个钱币一正一反,则局中人 1 和 2 分别得 $-1$ 和 1 分;或者说,局中人 1 输给局中人 2 一个单位.

可以用一个方阵来表示这些结果:

| | | 局中人 2 | |
|---|---|---|---|
| | | 1(正面) | 2(反面) |
| 局中人 1 | 1(正面) | 1 | $-1$ |
| | 2(反面) | $-1$ | 1 |

我们说,局中人 1 和 2 各有两个**策略**(strategy). 每个局中人的第 1 个策略表示选择出示钱币的正面,第 2 个策略表示选择反面.

这种游戏就是一个对策. 所谓对策,就是一组规则,它指明了整个游戏(或竞赛、或竞争、或斗争)自始至终所应遵循的各项规定和章程,包括局中人、策略、选定策略后的结局等.

上列表格称为对策的**支付矩阵**(payoff matrix),它是两个局中人的策略的函数. 例如,若局中人 1 出反面(策略 2),局中人 2 也出反面(策略 2),则在上表中第 2 行第 2 列处的元素 1 就是局中人 2 应该付给局中人 1 的数目(例如以元为单位);我们说,局中人 1 得到**支付**(payoff)1. 又如局中人 1 选择策略 1(正面),局中人

2 选择策略 2(反面),这时第 1 行第 2 列的元素是 $-1$,表示局中人 1 得到支付 $-1$,就是局中人 1 输掉 1 个单位;换句话说,局中人 2 从局中人 1 处赢进 1 个单位.

**例 6.1.2** "锤子、剪刀、布"游戏

每个小孩都玩过这种游戏.锤子胜剪刀,剪刀胜布,布胜锤子.这里也是两个局中人:局中人 1 和局中人 2.每人有 3 个策略:策略 1 代表出锤子,策略 2 代表出剪刀,策略 3 代表出布.假设胜者得 1 分,负者得 $-1$ 分,则支付矩阵是

| | | 局中人 2 | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| 局中人 1 | 1 | 0 | 1 | −1 |
| | 2 | −1 | 0 | 1 |
| | 3 | 1 | −1 | 0 |

**例 6.1.3** 局中人 1 从 $p=0,1,2,3$ 四个数中选出一个数,局中人 2 在不知道局中人 1 出什么数的情况下从 $q=0,1,2$ 三个数中选出一个数.局中人 1 得到的支付(即局中人 2 付给局中人 1 的数目)由下式决定:

$$P=q^2-p^2+2pq.$$

这是一个二人对策.局中人 1 有 4 个策略,局中人 2 有 3 个策略.支付矩阵如下:

| | | 局中人 2 | | |
|---|---|---|---|---|
| | $p$ \ $q$ | 0 | 1 | 2 |
| 局中人 1 | 0 | 0 | 1 | 4 |
| | 1 | −1 | 2 | 7 |
| | 2 | −4 | 1 | 8 |
| | 3 | −9 | −2 | 7 |

在以上三个例子里，都有 1,2 两个局中人，都有一个支付矩阵. 每个局中人有若干个策略；局中人 1 的策略用支付矩阵的行数来表示，局中人 2 的策略用支付矩阵的列数来表示. 每个局中人选定一个策略后，支付矩阵中有一个对应的元素，它代表局中人 2 应当付给局中人 1 的支付值. 正的支付值表示局中人 1 得到的支付是正值，这就是说，局中人 1 从局中人 2 处赢进若干个单位. 反之，负的支付值表示局中人 1 从局中人 2 处得到的是负值，这就是说，局中人 1 被局中人 2 赢得若干个单位.

### 6.1.2 矩阵对策

**定义 6.1.4** 设局中人 1 有 $m$ 个策略 $i=1,\cdots,m$；局中人 2 有 $n$ 个策略 $j=1,\cdots,n$. 设局中人 1,2 分别选择策略 $i$ 和 $j$ 时，局中人 1 从局中人 2 处得到的支付是 $a_{ij}$，则支付矩阵是

$$\boldsymbol{A}=(a_{ij})_{m\times n}.$$

对策由上列矩阵完全确定. 这种对策叫做**矩阵对策**(matrix game). $\{1,\cdots,m\}$和$\{1,\cdots,n\}$分别是局中人 1 和 2 的**策略集**(strategy set).

$a_{ij}$是局中人 1 得到的支付，局中人 2 得到的支付则是$-a_{ij}$. 这就是说，局中人 1 得到的就是局中人 2 付出的. 由于参加者只有两个，而且在任何情况下双方得到的支付之和为零，即 $a_{ij}+(-a_{ij})=0$，这种对策是零和二人对策.

局中人 1 希望支付越大越好，局中人 2 则相反，他希望支付 $a_{ij}$越小越好，即$(-a_{ij})$越大越好. 但是，在支付矩阵$(a_{ij})$中，每个局中人只能控制两个变量 $i$ 和 $j$ 中的一个：局中人 1 只能控制变量 $i$，局中人 2 只能控制变量 $j$.

如果局中人 1 选定一个策略 $i$，则他至少可以得到支付

$$\min_{1\leqslant j\leqslant n} a_{ij}.$$

这就是支付矩阵第 $i$ 行的最小元素. 由于局中人 1 希望支付值尽

可能地大，他可以选择 $i$ 使上式为最大. 即，他可以使支付不小于

$$\max_{1\leqslant i\leqslant m}\ \min_{1\leqslant j\leqslant n}\ a_{ij}. \tag{6.1}$$

同理，如果局中人 2 选定一个策略 $j$，则局中人 1 得到的支付至多是

$$\max_{1\leqslant i\leqslant m} a_{ij}.$$

这是支付矩阵第 $j$ 列的最大元素. 由于局中人 2 希望支付值尽可能地小，他可以选择 $j$ 使上式为最小，这就是说，他可以使局中人 1 得到的支付不大于

$$\min_{1\leqslant j\leqslant n}\ \max_{1\leqslant i\leqslant m}\ a_{ij}. \tag{6.2}$$

(6.1)就是支付矩阵各行最小值中的最大者，(6.2)则是其各列最大值中的最小者. 容易看出，在例 6.1.1 和例 6.1.2 的对策中，都有

$$-1=\max_{1\leqslant i\leqslant m}\ \min_{1\leqslant j\leqslant n} a_{ij}<\min_{1\leqslant j\leqslant n}\ \max_{1\leqslant i\leqslant m} a_{ij}=1;$$

而在例 6.1.3 的对策中，则有

$$\max_{1\leqslant i\leqslant m}\ \min_{1\leqslant j\leqslant n} a_{ij}=0=\min_{1\leqslant j\leqslant n}\ \max_{1\leqslant i\leqslant m} a_{ij}.$$

在一般情形，对于任意 $m\times n$ 矩阵对策

$$\boldsymbol{A}=(a_{ij}),$$

必有

$$\max_{1\leqslant i\leqslant m}\ \min_{1\leqslant j\leqslant n} a_{ij}\leqslant\min_{1\leqslant j\leqslant n}\ \max_{1\leqslant i\leqslant m} a_{ij}. \tag{6.3}$$

当上式中等号成立时，即

$$\max_{1\leqslant i\leqslant m}\ \min_{1\leqslant j\leqslant n} a_{ij}=\min_{1\leqslant j\leqslant n}\ \max_{1\leqslant i\leqslant m} a_{ij}=v \tag{6.4}$$

时，称这个值 $v$ 为对策的值(value). 此时，必有一个 $i=i^*$ 和一个 $j=j^*$，使

$$a_{ij^*}\leqslant a_{i^*j^*}=v\leqslant a_{i^*j}. \tag{6.5}$$

对于一切 $i$ 和一切 $j$ 成立. 这就是说，如果局中人 1 采用策略 $i^*$，则若局中人 2 选择 $j^*$ 以外的策略，支付值不可能小于 $v$；如果局中

人 2 采用策略 $j^*$，则若局中人 1 选择 $i^*$ 以外的策略，支付值不可能大于 $v$.

$i^*$ 和 $j^*$ 分别称为局中人 1 和 2 的**最优策略**(optimal strategy). $(i^*,j^*)$是对策的一个**鞍点**(saddle point)，或称为对策的一个**解**(solution).

**定理 6.1.5** 如果$(i^*,j^*)$和$(i^\circ,j^\circ)$都是 $m\times n$ 矩阵对策 $\boldsymbol{A}=(a_{ij})$的鞍点，则$(i^*,j^\circ)$和$(i^\circ,j^*)$也都是它的鞍点，并且

$$a_{i^*j^*}=a_{i^\circ j^\circ}=a_{i^*j^\circ}=a_{i^\circ j^*}.$$

这个定理说明了若矩阵对策有鞍点，且鞍点不止一个，则它具有两个性质：一是鞍点的可交换性，或称为**矩形性质**(rectangular property)，二是鞍点处的值都相等.

### 6.1.3 混合策略

若矩阵对策 $\boldsymbol{A}=(a_{ij})$没有鞍点，即当

$$\max_{1\leqslant i\leqslant m}\min_{1\leqslant j\leqslant n}a_{ij}<\min_{1\leqslant j\leqslant n}\max_{1\leqslant i\leqslant m}a_{ij}$$

时，6.1.2 中所述的值和解不存在. 现引入混合策略的概念.

为了避免让对方猜出自己采用哪一个策略，局中人 1 可以用一种随机的方法来决定自己要选择的策略，也就是采用一个混合策略.

**定义 6.1.6** 局中人 1 的**混合策略**(mixed strategy)是满足

$$x_i\geqslant 0,\ i=1,\cdots,m\quad 和\quad \sum_{i=1}^{m}x_i=1 \text{ 的 } m \text{ 维向量}$$

$$\boldsymbol{X}=(x_1,\cdots,x_m).$$

以 $S_m$ 记全体 $\boldsymbol{X}$ 的集. 同样，局中人 2 的混合策略是满足 $y_j\geqslant 0$，$j=1,\cdots,n$和 $\sum_{j=1}^{n}y_j=1$ 的 $n$ 维向量

$$\boldsymbol{Y}=(y_1,\cdots,y_n).$$

以 $S_n$ 记全体 $\boldsymbol{Y}$ 的集.

局中人 1 以概率 $x_i$ 选择策略 $i$，局中人 2 以概率 $y_j$ 选择策略 $j$. 这时，支付为 $a_{ij}$的概率是 $x_iy_j$. 每一个支付 $a_{ij}$乘以相应的概率 $x_iy_j$，对所有的 $i$ 和所有的 $j$ 求和，就得到局中人 1 的**期望支付**(expected payoff)

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_iy_j = \boldsymbol{XAY}^{\mathrm{T}}.$$

局中人 1 希望这个期望支付越大越好，局中人 2 则希望它越小越好. 如果局中人 1 选用策略 $\boldsymbol{X}\in S_m$，他的期望支付至少有

$$\min_{\boldsymbol{Y}\in S_n} \boldsymbol{XAY}^{\mathrm{T}}.$$

局中人 1 可以选择 $\boldsymbol{X}\in S_m$ 使上式为最大. 这就是说，他可以使自己得到的期望支付不小于

$$v_1 = \max_{\boldsymbol{X}\in S_m}\ \min_{\boldsymbol{Y}\in S_n} \boldsymbol{XAY}^{\mathrm{T}}. \tag{6.6}$$

如果局中人 2 选用策略 $\boldsymbol{Y}\in S_n$，他应当付出的期望支付至多是

$$\max_{\boldsymbol{X}\in S_m} \boldsymbol{XAY}^{\mathrm{T}}.$$

局中人 2 可以选择 $\boldsymbol{Y}\in S_n$ 使上式为最小. 这就是说，他可以使局中人 1 得到的期望支付不大于

$$v_2 = \min_{\boldsymbol{Y}\in S_n}\ \max_{\boldsymbol{X}\in S_m} \boldsymbol{XAY}^{\mathrm{T}}. \tag{6.7}$$

J. von Neumann 首先证明了对于一切 $m\times n$ 矩阵对策 $\boldsymbol{A}=(a_{ij})$，值 $v_1$ 和 $v_2$ 存在且相等. 这一结果就是著名的对策论基本定理，或者叫做**最小最大值定理**(minimax theorem).

**定理 6.1.7** (von Neumann 定理) 对于任意 $m\times n$ 矩阵对策 $\boldsymbol{A}=(a_{ij})$，

$$\max_{\boldsymbol{X}\in S_m}\ \min_{\boldsymbol{Y}\in S_n} \boldsymbol{XAY}^{\mathrm{T}} = \min_{\boldsymbol{Y}\in S_n}\ \max_{\boldsymbol{X}\in S_m} \boldsymbol{XAY}^{\mathrm{T}}.$$

(6.5)是**纯策略**(pure strategy)下鞍点的定义. 对于混合策略，也有类似的鞍点的概念.

**定义 6.1.8** 设 $X^*\in S_m, Y^*\in S_n$. 如果对于一切 $X\in S_m$ 和一切 $Y\in S_n$ 有

$$XAY^{*T}\leqslant X^*AY^{*T}\leqslant X^*AY^T,$$

则称 $(X^*,Y^*)$ 是 $m\times n$ 矩阵对策 $A=(a_{ij})$ 的一个**鞍点**(在混合策略意义下).

下面的定理表明鞍点存在和最小最大值定理的等价性.

**定理 6.1.9** $m\times n$ 矩阵对策 $A=(a_{ij})$ 有鞍点的充要条件是 (6.6) 和 (6.7) 存在且相等.

我们将定理 6.1.7 中等式的左右两边改写成下列等价形式:

$$\max_{X\in S_m}\ \min_{Y\in S_n}\sum_{j=1}^{n}\Big(\sum_{i=1}^{m}a_{ij}x_i\Big)y_j$$

$$=\max_{X\in S_m}\ \min_{1\leqslant j\leqslant n}\sum_{i=1}^{m}a_{ij}x_i,$$

$$\min_{Y\in S_n}\ \max_{X\in S_m}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}a_{ij}y_j\Big)x_i$$

$$=\min_{Y\in S_n}\ \max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}a_{ij}y_j,$$

于是最小最大值定理 6.1.7 就可以写成下面较简单实用的形式了:

$$\max_{X\in S_m}\ \min_{1\leqslant j\leqslant n}\sum_{i=1}^{m}a_{ij}x_i$$

$$=\min_{Y\in S_n}\ \max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}a_{ij}y_j. \tag{6.8}$$

### 6.1.4 最优策略及其性质

设 $m\times n$ 矩阵对策的支付矩阵是 $A=(a_{ij})$. 由定理 6.1.7 和定理 6.1.9 可知,鞍点必定存在. 设 $(X^*,Y^*)$ 是一个鞍点,即

$$XAY^{*T}\leqslant X^*AY^{*T}\leqslant X^*AY^T$$

对于一切$X \in S_m$和一切$Y \in S_n$成立.

我们称$X^*$和$Y^*$分别是局中人1和2的最优策略，并称$X^* A Y^{*T}$是对策的值. 这就是说，在鞍点$(X^*, Y^*)$处，对策的期望支付就是对策的值. 也称$(X^*, Y^*)$是对策的一个解. 由定理6.1.9可知，对策的值就是

$$v_1 = \max_{X \in S_m} \min_{Y \in S_n} XAY^{\mathrm{T}}$$

和

$$v_2 = \min_{Y \in S_n} \max_{X \in S_m} XAY^{\mathrm{T}}$$

的共同值$v$.

鞍点的定义

$$XAY^{*T} \leqslant X^* A Y^{*T} \leqslant X^* A Y^{\mathrm{T}}$$

表明，只要局中人1坚持采用他的最优策略$X^*$，则不论局中人2选择什么策略，局中人1的期望支付不会少于$X^* A Y^{*T}$，这个$X^* A Y^{*T}$就是对策的值；同样，只要局中人2坚守他的最优策略$Y^*$，则不论局中人1选择什么策略，都不能使期望支付超过$X^* A Y^{*T}$.

我们采用一些常见的矩阵记号.

以$A_{i\cdot}$表示$m \times n$矩阵$A$的第$i$个行向量，以$A_{\cdot j}$表示$A$的第$j$个列向量，则

$$XA_{\cdot j} = \sum_{i=1}^{m} a_{ij} x_i ,$$

$$A_{i\cdot} Y^{T} = \sum_{j=1}^{n} a_{ij} y_j .$$

第一个式子表示局中人1采用混合策略$X$而局中人2采用纯策略$j$时的期望支付；第二个式子表示局中人1采用纯策略$i$而局中人2采用混合策略$Y$时的期望支付.

下面是最优策略的两个最主要的性质.

**定理 6.1.10** 设 $m\times n$ 矩阵对策 $A=(a_{ij})$ 的值是 $v$.

(1) 设 $Y^*$ 是局中人 2 的一个最优策略. 如果对于某个 $i$ 有 $A_{i\cdot}Y^{*T}<v$,则在局中人 1 的任一个最优策略 $X^*$ 中必有 $x_i^*=0$.

(2) 设 $X^*$ 是局中人 1 的一个最优策略. 如果对于某个 $j$ 有 $X^*A_{\cdot j}>v$,则在局中人 2 的任一个最优策略 $Y^*$ 中必有 $y_j^*=0$.

这个定理的(1)告诉我们,如果已知对策的值是 $v$,并且 $Y^*$ 是局中人 2 的最优策略,若局中人 1 采用纯策略 $i$ 时他的期望支付达不到 $v$,则纯策略 $i$ 是不可取的,在局中人 1 的任何一个最优策略 $X^*$ 中一定不会包含这个纯策略. 换一种说法,即:如果已知局中人 1 的某个最优策略 $X^*$ 中有 $x_i^*>0$,则必有 $A_{i\cdot}Y^{*T}=v$.

定理中(2)的意义与此类似.

**定理 6.1.11** 设 $m\times n$ 矩阵对策 $A=(a_{ij})$ 的值是 $v$.

(1) $X^*\in S_m$ 是局中人 1 的最优策略的充要条件是

$$X^*A_{\cdot j}\geqslant v,\quad j=1,\cdots,n.$$

(2) $Y^*\in S_n$ 是局中人 2 的最优策略的充要条件是

$$A_{i\cdot}Y^{*T}\leqslant v,\quad i=1,\cdots,m.$$

如果已知对策的值,可以利用这个定理检验局中人 1 或 2 的某个策略是否是他的最优策略.

### 6.1.5 策略的优超关系

考虑 $3\times 3$ 矩阵对策

$$\begin{bmatrix}0 & 1 & -1\\ 1 & -2 & 0\\ 2 & -1 & 1\end{bmatrix}.$$

对支付矩阵的元素稍加考察,不难看出,局中人 1 决不会采用他的第 2 个策略. 这是因为,不论局中人 2 选择什么策略,局中人 1 的第 3 个策略的支付总比第 2 个策略的支付为大. 因此,局中人 1 的第 2 个策略只能以零概率出现在他的最优混合策略里.

于是，要解上面的矩阵对策，可以将矩阵的第 2 行划去．只要解矩阵对策

$$\begin{bmatrix} 0 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix}$$

就行了．

对于这个 $2\times3$ 矩阵对策，局中人 2 显然不愿采用策略 1．不论局中人 1 选用哪一个策略，局中人 2 的第 3 个策略的支付都小于第 1 个策略的支付．

因此，可以将上面这个矩阵的第 1 列划去，只要解矩阵对策

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

就行了．而这个 $2\times2$ 矩阵对策，容易验证（利用定理 6.1.11），它的值是 $v=0$，局中人 1 和 2 的最优策略分别是 $\boldsymbol{X}^*=(1/2,1/2)$，$\boldsymbol{Y}^*=(1/2,1/2)$．关于 $2\times2$ 矩阵对策的解法，见后面的 6.1.6．

回到原来的 $3\times3$ 矩阵对策，显然它的解应是 $\boldsymbol{X}^*=(1/2,0,1/2)$，$\boldsymbol{Y}^*=(0,1/2,1/2)$，$v=0$．

由此可见，在原来的 $3\times3$ 矩阵对策中，局中人 1 的第 3 个策略比第 2 个策略好，得到的支付大，所以他不必考虑策略 2．在划去第 2 行后的 $2\times3$ 矩阵对策中，局中人 2 的第 3 个策略比第 1 个策略好，付出的支付小，所以他不会采用策略 1．

下面是策略的**优超关系**（domination）的定义．

**定义 6.1.12** 设 $A=(a_{ij})$ 是 $m\times n$ 矩阵对策．如果 $a_{kj}\geqslant a_{lj}$，$j=1,\cdots,n$，则称局中人 1 的策略 $k$ 优超于策略 $l$．

如果 $a_{ik}\leqslant a_{il}$，$i=1,\cdots,m$，则称局中人 2 的策略 $k$ 优超于策略 $l$．

如果在以上两组式子里严格的不等号成立，则分别称局中人 1，2 的策略 $k$ **严格优超**于策略 $l$（strategy $k$ strictly dominates strategy $l$）．

对于混合策略，也有类似的优超关系概念.

现在只考虑一个纯策略被另外几个纯策略的凸线性组合所优超的情形.

可以证明，在这种情形，如果是严格优超，则将被优超的那个纯策略所对应的行或列划去后，从余下的较小的矩阵对策的最优策略，就立即可以得到原来对策的最优策略，这只要将划去的那一行或列所对应的纯策略赋以概率零就行了.

如果是优超而不是严格优超，仍可以从划去某些行或列后的矩阵对策的解得到原来对策的解. 但这时有可能"失去"某些解. 这就是说，从较小矩阵对策的最优策略，通过加上零概率的办法得到原来对策的最优策略，所得到的可能不是全部解. 然而，在通常的情形，往往只要求得到一个解，而不是全部解，这时就可以应用这种优超关系来简化求解过程.

**例 6.1.13** 设矩阵对策的支付矩阵是

$$\begin{bmatrix} 0 & 3 & -2 & 4 \\ 9 & 0 & 4 & -1 \\ 3 & 3 & 4 & 2 \\ 1 & 4 & 3 & 4 \end{bmatrix}.$$

局中人 1 的策略 1 被他的策略 4 优超，因此可将矩阵的第 1 行划去：

$$\begin{bmatrix} 9 & 0 & 4 & -1 \\ 3 & 3 & 4 & 2 \\ 1 & 4 & 3 & 4 \end{bmatrix}.$$

局中人 2 的第 2 个策略被他的第 4 个策略优超，因此可将上面这个矩阵的第 2 列划去：

$$\begin{bmatrix} 9 & 4 & -1 \\ 3 & 4 & 2 \\ 1 & 3 & 4 \end{bmatrix}.$$

不难看出，这个 3×3 矩阵的元素满足下列关系：

$$\begin{bmatrix}4\\4\\3\end{bmatrix} \geqslant \frac{1}{3}\begin{bmatrix}9\\3\\1\end{bmatrix} + \frac{2}{3}\begin{bmatrix}-1\\2\\4\end{bmatrix}.$$

因此，又可以将这个 3×3 矩阵的第 2 列划去，得到

$$\begin{bmatrix}9 & -1\\3 & 2\\1 & 4\end{bmatrix}.$$

这个 3×2 矩阵的第 2 行元素被第 1 和第 3 行元素的一个凸线性组合所优超：

$$(3,2) \leqslant \frac{1}{4}(9,-1) + \frac{3}{4}(1,4).$$

因此，又可以将这个 3×2 矩阵的第 2 行划去，得到

$$\begin{bmatrix}9 & -1\\1 & 4\end{bmatrix}.$$

容易验证，这个 2×2 矩阵对策的最优策略是 $X^* = \left(\frac{3}{13}, \frac{10}{13}\right)$，$Y^* = \left(\frac{5}{13}, \frac{8}{13}\right)$，值为 $v = 37/13$.

这个 2×2 矩阵是原来的 4×4 矩阵第 2，4 行和第 1，4 列组成的子矩阵. 因此，

$$X^* = \left(0, \frac{3}{13}, 0, \frac{10}{13}\right), Y^* = \left(\frac{5}{13}, 0, 0, \frac{8}{13}\right)$$

是原来的矩阵对策的最优策略；它的值也是 $v = 37/13$.

**例 6.1.14** 设矩阵对策的支付矩阵是

$$\begin{bmatrix}3 & 3 & 5\\4 & 2 & -3\\3 & 3 & 2\end{bmatrix}.$$

这个对策有一个鞍点，在 $i=1, j=2$ 处，因为元素 $a_{12}=3$ 是它

所在的行中的最小元素，同时是它所在的列中的最大元素. 因此，(1,2)是对策的解，$a_{12}=3$ 是对策的值.

如果利用策略的优超关系就可以先划去矩阵的第 3 行，然后在剩下的 2×3 矩阵中再划去第 1 列. 余下一个 2×2 矩阵

$$\begin{bmatrix} 3 & 5 \\ 2 & -3 \end{bmatrix}.$$

这里左上角元素 3 所在的位置仍是一个鞍点. 对于原来的 3×3 矩阵来说，仍然得到最优策略

$$\boldsymbol{X}_1^* = (1,0,0), \quad \boldsymbol{Y}_1^* = (0,1,0).$$

但是，除此以外，混合策略

$$\boldsymbol{X}_2^* = \left(\frac{1}{3},0,\frac{2}{3}\right), \quad \boldsymbol{Y}_2^* = \left(\frac{1}{2},\frac{1}{2},0\right)$$

也是对策的一个解. 这一事实可以应用定理 6.1.11 得到验证. 参看后面的例 6.1.15.

上例表明，利用策略的优超(不是严格优超)关系划去支付矩阵的某些行或列得出较小的矩阵对策的解，加上零概率后成为原来对策的解，在这一过程中可能会"失去"原来对策的一些解.

### 6.1.6　2×2 矩阵对策的解

设 2×2 矩阵对策的支付矩阵是

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{6.9}$$

如果对策有纯策略的鞍点，立即得到纯策略解.

在没有鞍点的情况下，通过两行互换或两列互换，也就是通过局中人 1 或 2 的两个策略编号的互换，或者通过矩阵的转置和各元素的变号，也就是通过局中人 1,2 名称的互换，不难发现，只需考虑下列情形：

$$a < b, \quad a < c, \quad d < b, \quad d < c.$$

这时，对策必有混合策略解。

设$X^* = (x_1^*, x_2^*) = (x^*, 1 - x^*)$，

$Y^* = (y_1^*, y_2^*) = (y^*, 1 - y^*)$

分别是局中人 1 和 2 的最优混合策略，其中

$$0 < x^* < 1, \quad 0 < y^* < 1.$$

根据定理 6.1.10，由于

$$x^* > 0, \quad 1 - x^* > 0, \quad y^* > 0, \quad 1 - y^* > 0,$$

因此，如果以 $v$ 表示对策的值，必有

$$\boldsymbol{X}^* \boldsymbol{A}_{\cdot 1} = v, \quad \boldsymbol{X}^* \boldsymbol{A}_{\cdot 2} = v,$$

$$\boldsymbol{A}_{1\cdot} \boldsymbol{Y}^{*\mathrm{T}} = v, \quad \boldsymbol{A}_{2\cdot} \boldsymbol{Y}^{*\mathrm{T}} = v.$$

将这 4 个方程用(6.9)的元素写出来，就是

$$ax^* + c(1 - x^*) = v,$$

$$bx^* + d(1 - x^*) = v,$$

$$ay^* + b(1 - y^*) = v,$$

$$cy^* + d(1 - y^*) = v.$$

由前两式可得

$$x^* = \frac{d - c}{a + d - b - c}, \tag{6.10}$$

由后两式可得

$$y^* = \frac{d - b}{a + d - b - c}. \tag{6.11}$$

然后就可求得

$$v = \frac{ad - bc}{a + d - b - c}. \tag{6.12}$$

(6.10)～(6.12)就是 2×2 矩阵对策(6.9)当它不存在鞍点时的最优策略和值.

这些公式对于

$$a > b, \quad a > c, \quad d > b, \quad d > c$$

的情形同样适用.

### 6.1.7 2×n 和 m×2 矩阵对策的图解法

对于 2×n 的情形，以 $n=3$ 为例，设 2×3 矩阵对策的支付矩阵 **A** 为

$$\begin{array}{ccc} & & \boxed{1} \quad \boxed{2} \quad \boxed{3} \\ x & ① & \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}. \\ 1-x & ② & \end{array}$$

以①，②分别表示局中人 1 的第 1，2 个纯策略；$\boxed{1}$，$\boxed{2}$，$\boxed{3}$表示局中人 2 的第 1，2，3 个纯策略.

设局中人 1 采用混合策略 $\boldsymbol{X}=(x,1-x)$，这里 $0\leqslant x\leqslant 1$. $x=1$ 代表纯策略①，$x=0$ 代表纯策略②.

当 $x=1$，即局中人 1 采用纯策略①时，若局中人 2 采用纯策略$\boxed{1}$，支付为 $a$，见图 6.1. 当 $x=0$，即局中人 1 采用纯策略②时，对应于$\boxed{1}$的支付为 $d$. 连接图中的直线 $ad$.

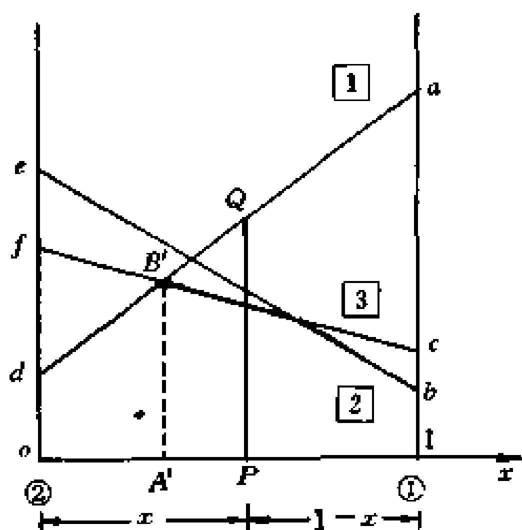



图 6.1

设 $x$ 轴上点 $P$ 的坐标是 $x$. 容易证明，纵坐标 $PQ$ 是局中人 1 采用混合策略 $\boldsymbol{X}$ 而局中人 2 采用纯策略 $\boxed{1}$ 时的期望支付，即 $\boldsymbol{X}A_{\cdot 1}$.

同样，$be$ 和 $cf$ 上点的纵坐标分别表示局中人 1 采用 $\boldsymbol{X}$ 而局中人 2 采用纯策略 $\boxed{2}$ 和 $\boxed{3}$ 时的期望支付.

对于局中人 1 的每一个混合策略，他至少可以得到 $ad, be, cf$ 三条直线在 $x$ 处纵坐标的最小值，即

$$\min_{1\leqslant j\leqslant 3} \boldsymbol{X}\boldsymbol{A}_{\cdot j} = \min_{1\leqslant j\leqslant 3} \sum_{i=1}^{2} a_{ij}x_i.$$

图 6.1 中的粗折线表示这个最小值函数.

局中人 1 希望选择 $\boldsymbol{X}$ 使上面这个最小值尽可能地大. 从图中可以看出来，他应当选择点 $A'$ 所代表的 $\boldsymbol{X}$，这时上述最小值为最大，即

$$A'B' = \max_{\boldsymbol{X}\in S_2} \min_{1\leqslant j\leqslant 3} \boldsymbol{X}\boldsymbol{A}_{\cdot j}.$$

由式(6.8)可知，这就是对策的值.

在图 6.1 中，点 $B'$ 是 $ad$ 和 $cf$ 两条直线的交点. 只要解两个二元一次联立方程，就可求出 $A'$ 的坐标 $x=x^*$ 和 $A'B'$ 的值. 由图 6.1 也可看出，局中人 2 的最优策略不涉及他的纯策略 $\boxed{2}$. 因此，只要解 $2\times 2$ 矩阵对策

$$\begin{bmatrix} a & c \\ d & f \end{bmatrix},$$

就能得出局中人 2 的最优策略.

这种图解法可以推广到一切 $2\times n$ 的矩阵对策.

在特殊情形，得到的解可以是 $x$ 轴上 $[0,1]$ 的一个子区间，也可以是 $[0,1]$ 的一个端点. 前者是图 6.1 中的粗折线含有一段水平线段的情形，后者则对应于局中人 1 的一个纯策略解.

下面再来讨论 $m\times 2$ 矩阵对策的图解法. 也以 $m=3$ 的情形来加以说明.

设 3×2 矩阵对策的支付矩阵 $\boldsymbol{A}$ 为

$$\begin{array}{cc} & \begin{array}{cc} y & 1-y \\ \boxed{1} & \boxed{2} \end{array} \\ \begin{array}{c} ① \\ ② \\ ③ \end{array} & \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}. \end{array}$$

设局中人 2 采用混合策略 $\boldsymbol{Y}=(y,1-y),0\leqslant y\leqslant 1$. $y=1$ 代表纯策略$\boxed{1}$, $y=0$ 代表纯策略$\boxed{2}$.

图 6.2 中粗折线的纵坐标是

$$\max_{1\leqslant i\leqslant 3}\boldsymbol{A}_{i\cdot}\boldsymbol{Y}^{\mathrm{T}}=\max_{1\leqslant i\leqslant 3}\sum_{j=1}^{2}a_{ij}y_j.$$

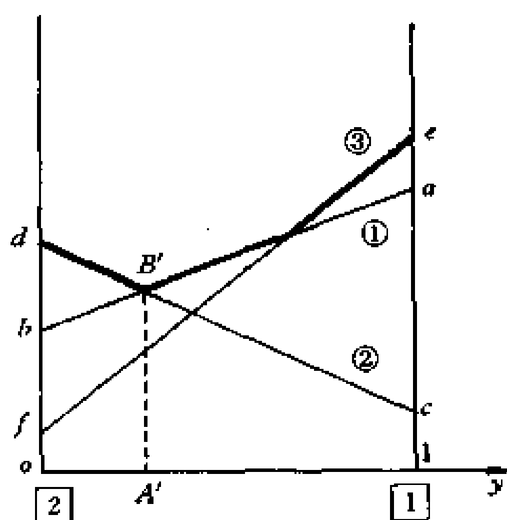



图 6.2

局中人 2 希望选择使上面这个最大值尽可能小的 $\boldsymbol{Y}$. 在图上就是点 $A'$ 所代表的 $\boldsymbol{Y}$. 这时

$$A'B'=\min_{\boldsymbol{Y}\in S_2}\ \max_{1\leqslant i\leqslant 3}\boldsymbol{A}_{i\cdot}\boldsymbol{Y}^{\mathrm{T}}.$$

这就是对策的值.

### 6.1.8 3×3 矩阵对策的解

$S_3$ 中的点 $\boldsymbol{X}=(x_1,x_2,x_3)$ 满足条件

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0, \tag{6.13}$$

$$x_1 + x_2 + x_3 = 1. \tag{6.14}$$

可以用下述方法在平面上表示这些点.

设 $\boldsymbol{X}$ 是高为 1 的等边三角形 123(图 6.3)中的任意一点.如果从点 $\boldsymbol{X}$ 到 1,2,3 这三个顶点的对边的距离分别为 $x_1,x_2,x_3$,则 $x_1$, $x_2$, $x_3$ 满足(6.13)和(6.14);因此,就以$(x_1,x_2,x_3)$作为点 $X$ 的坐标,称为**重心坐标**(barycentric coordinates).顶点 1 的重心坐标是(1,0,0).类似地,顶点 2,3 的重心坐标分别是(0,1,0)和(0,0,1).闭三角形中全部点的集合就是 $S_3$.三角形的三条边 23,31,12 的方程分别是 $x_1=0,x_2=0,x_3=0$.

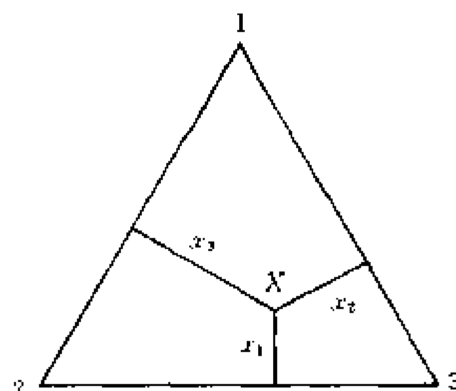


图 6.3

设 3×3 矩阵对策的支付矩阵是

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{6.15}$$

根据(6.8),对策的值是

$$\begin{aligned} v &= \max_{X\in S_3} \min_{1\leqslant j\leqslant 3} \boldsymbol{XA}_{\cdot j} \\ &= \max_{X\in S_3} \min \{\boldsymbol{XA}_{\cdot 1},\boldsymbol{XA}_{\cdot 2},\boldsymbol{XA}_{\cdot 3}\}. \end{aligned} \tag{6.16}$$

考虑方程

$$\boldsymbol{XA}_{\cdot 1} = \boldsymbol{XA}_{\cdot 2}, \tag{6.17}$$

$$XA_{\cdot 2} = XA_{\cdot 3}, \qquad (6.18)$$

$$XA_{\cdot 3} = XA_{\cdot 1}. \qquad (6.19)$$

每一个方程代表一条直线，它将整个平面分成两个半平面（可以把三角形外面的点看做是满足条件(6.14)而 $x_1, x_2, x_3$ 三数中有一个或两个取负值的点）. 例如方程(6.17)，它将整个平面分成两个半平面. 一个半平面中的点 $X$ 满足条件 $XA_{\cdot 1} < XA_{\cdot 2}$，另一个半平面中的点 $X$ 满足条件 $XA_{\cdot 1} > XA_{\cdot 2}$.

对于方程(6.18)和(6.19)，情况也一样.

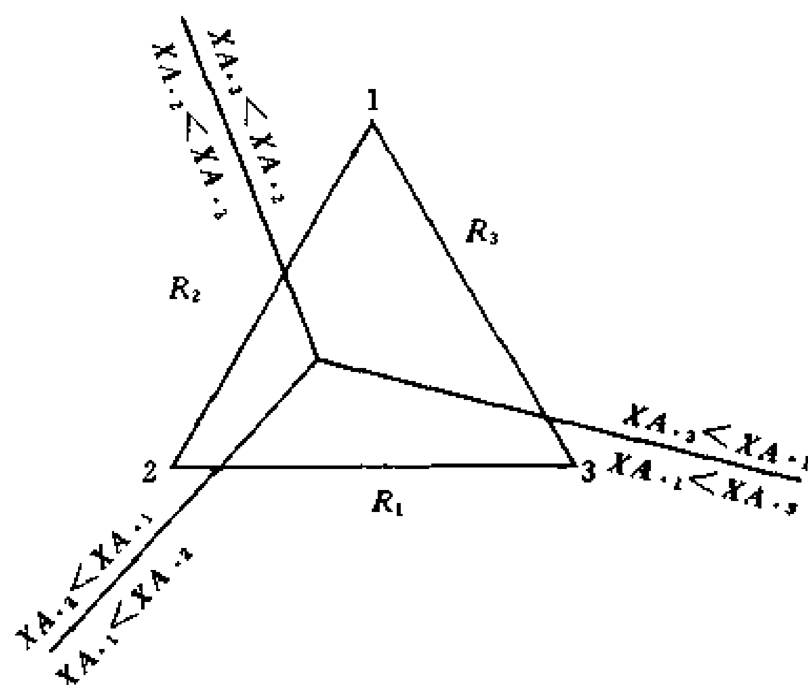



图 6.4

三条直线(6.17)～(6.19)或交于一点，或互相平行，它们将整个平面分为三个区域 $R_1, R_2, R_3$，见图 6.4. 这里

$$R_1 = \{X \mid X \in S_3;\ \min_{1 \leqslant j \leqslant 3} XA_{\cdot j} = XA_{\cdot 1}\},$$

$$R_2 = \{X \mid X \in S_3;\ \min_{1 \leqslant j \leqslant 3} XA_{\cdot j} = XA_{\cdot 2}\},$$

$$R_3 = \{X \mid X \in S_3;\ \min_{1 \leqslant j \leqslant 3} XA_{\cdot j} = XA_{\cdot 3}\}.$$

因此,(6.16)可以改写成

$$
\begin{aligned}
v &= \max_{X\in S_3}\min_{1\leqslant j\leqslant 3} XA_{\cdot j} \\
&= \max\{\max_{X\in S_3\cap R_1} XA_{\cdot 1},\ \max_{X\in S_3\cap R_2} XA_{\cdot 2},\ \max_{X\in S_3\cap R_3} XA_{\cdot 3}\}.
\end{aligned}
$$

应当先计算

$$
\max_{X\in S_3\cap R_j} XA_{\cdot j},\quad j=1,2,3.
$$

然后在这些值中取其最大者即为 $v$.

由于 $S_3\cap R_j(j=1,2,3)$是凸多边形(在特殊情况可以是一个直线段、或一个点、或空集),一次函数 $XA_{\cdot j}$在凸多边形域上只可能在顶点处取得其最大值,因此,只需计算 $XA_{\cdot j}$在有关顶点处的值,进行比较,其最大者就是 $v$. 在比较的过程中,同时也就得到了局中人 1 的最优策略.

求出对策的值 $v$ 以后,不难用类似的方法计算局中人 2 的最优策略 $\boldsymbol{Y}$. 我们有

$$
\begin{aligned}
v &= \min_{Y\in S_3}\max_{1\leqslant i\leqslant 3} A_{i\cdot}Y^{\mathrm{T}} \\
&= \min_{Y\in S_3}\max\ \{A_{1\cdot}Y^{\mathrm{T}}, A_{2\cdot}Y^{\mathrm{T}}, A_{3\cdot}Y^{\mathrm{T}}\} \\
&= \min\{\min_{Y\in S_3\cap T_1} A_{1\cdot}Y^{\mathrm{T}},\ \min_{Y\in S_3\cap T_2} A_{2\cdot}Y^{\mathrm{T}},\ \min_{Y\in S_3\cap T_3} A_{3\cdot}Y^{\mathrm{T}}\},
\end{aligned}
$$

其中

$$
\begin{aligned}
T_1 &= \{Y\mid Y\in S_3;\max_{1\leqslant i\leqslant 3} A_{i\cdot}Y^{\mathrm{T}} = A_{1\cdot}Y^{\mathrm{T}}\}, \\
T_2 &= \{Y\mid Y\in S_3;\max_{1\leqslant i\leqslant 3} A_{i\cdot}Y^{\mathrm{T}} = A_{2\cdot}Y^{\mathrm{T}}\}, \\
T_3 &= \{Y\mid Y\in S_3;\max_{1\leqslant i\leqslant 3} A_{i\cdot}Y^{\mathrm{T}} = A_{3\cdot}Y^{\mathrm{T}}\}.
\end{aligned}
$$

只须计算 $A_{i\cdot}Y^{\mathrm{T}}$ 在有关顶点处的值,进行比较,其最小者必为 $v$. 而取得最小值 $v$ 的点 $\boldsymbol{Y}$ 就是局中人 2 的最优策略.

**例 6.1.15** 应用上述方法计算例 6.1.14 中对策的值和最优策略. 支付矩阵是

$$B = \begin{bmatrix} 3 & 3 & 5 \\ 4 & 2 & -3 \\ 3 & 3 & 2 \end{bmatrix}.$$

为了计算简便，对矩阵 $\boldsymbol{B}$ 的每个元素加上一个常数 $-3$，得到

$$A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & -1 & -6 \\ 0 & 0 & -1 \end{bmatrix}.$$

现在来解矩阵对策 $\boldsymbol{A}$. 有

$$XA_1 = x_2,$$
$$XA_2 = -x_2,$$
$$XA_3 = 2x_1 - 6x_2 - x_3.$$

直线 $XA_{\cdot 1} = XA_{\cdot 2}$的方程是 $x_2 = 0$；直线 $XA_{\cdot 2} = XA_{\cdot 3}$的方程是

$$2x_1 - 5x_2 - x_3 = 0,$$

或

$$3x_1 - 4x_2 = 1;$$

直线 $XA_{\cdot 3} = XA_{\cdot 1}$的方程是

$$2x_1 - 7x_2 - x_3 = 0,$$

或

$$3x_1 - 6x_2 = 1.$$

在图 6.5 中画出了 $\min\limits_{1 \leqslant j \leqslant 3} XA_{\cdot j}$分别等于 $XA_{\cdot 1}, XA_{\cdot 2}, XA_{\cdot 3}$的三个区域 $R_1, R_2, R_3$.

计算 $XA_{\cdot 1}$在$(1,0,0)$和$\left(\frac{1}{3},0,\frac{2}{3}\right)$处的值：

$$(1,0,0)(0,1,0)^{\mathrm{T}} = 0,$$
$$\left(\frac{1}{3},0,\frac{2}{3}\right)(0,1,0)^{\mathrm{T}} = 0.$$

再计算 $XA_{\cdot 3}$在$(0,1,0)$，$(0,0,1)$和$\left(\frac{5}{7},\frac{2}{7},0\right)$处的值：

$$(0,1,0)(2,-6,-1)^{\mathrm{T}} = -6,$$
$$(0,0,1)(2,-6,-1)^{\mathrm{T}} = -1,$$

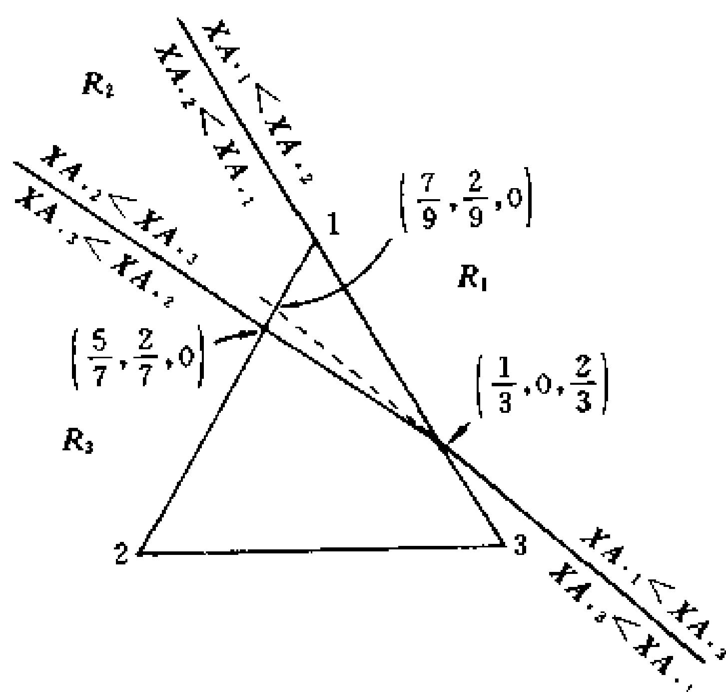



图 6.5

$$\left(\frac{5}{7},\frac{2}{7},0\right)(2,-6,-1)^{\mathrm{T}}=-\frac{2}{7}.$$

比较以上 5 个值,其最大者为 0. 因此,矩阵对策 $\boldsymbol{A}$ 的值是

$$v_A = 0. \tag{6.20}$$

局中人 1 的最优策略是

$$\boldsymbol{X}_1^* = (1,0,0), \quad \boldsymbol{X}_2^* = \left(\frac{1}{3},0,\frac{2}{3}\right). \tag{6.21}$$

以下再来计算局中人 2 的最优策略. 我们有

$$\boldsymbol{A}_{1\cdot}\boldsymbol{Y}^{\mathrm{T}} = 2y_3,$$
$$\boldsymbol{A}_{2\cdot}\boldsymbol{Y}^{\mathrm{T}} = y_1 - y_2 - 6y_3,$$
$$\boldsymbol{A}_{3\cdot}\boldsymbol{Y}^{\mathrm{T}} = -y_3.$$

直线 $\boldsymbol{A}_{1\cdot}\boldsymbol{Y}^{\mathrm{T}}=\boldsymbol{A}_{2\cdot}\boldsymbol{Y}^{\mathrm{T}}$ 的方程是

$$y_1 - y_2 - 8y_3 = 0, \quad 或 \quad 9y_1 + 7y_2 = 8;$$

直线 $\boldsymbol{A}_{2\cdot}\boldsymbol{Y}^{\mathrm{T}}=\boldsymbol{A}_{3\cdot}\boldsymbol{Y}^{\mathrm{T}}$ 的方程是

$$y_1 - y_2 - 5y_3 = 0, \quad 或 \quad 6y_1 + 4y_2 = 5;$$

直线 $A_3.Y^T=A_1.Y^T$ 的方程是

$$y_3=0.$$

在图 6.6 中画出了 $\max\limits_{1\leqslant i\leqslant 3}A_i.Y^T$ 分别等于 $A_1.Y^T, A_2.Y^T, A_3.Y^T$ 的 3 个区域 $T_1, T_2, T_3$.

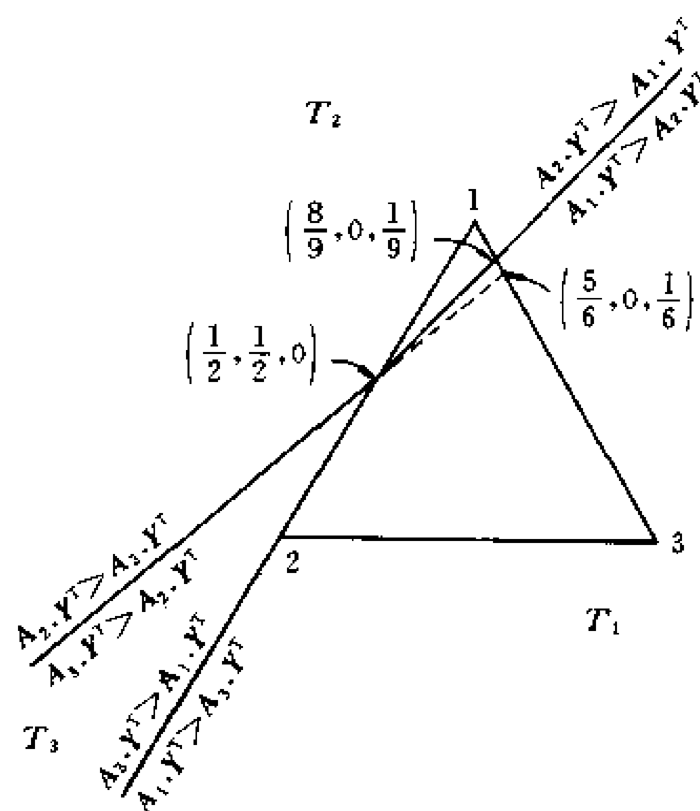



图 6.6

计算 $A_1.Y^T$ 在 $(0,1,0),(0,0,1),\left(\frac{8}{9},0,\frac{1}{9}\right),\left(\frac{1}{2},\frac{1}{2},0\right)$ 处的值：

$$(0,0,2)(0,1,0)^T=0,$$
$$(0,0,2)(0,0,1)^T=2,$$
$$(0,0,2)\left(\frac{8}{9},0,\frac{1}{9}\right)^T=\frac{2}{9},$$
$$(0,0,2)\left(\frac{1}{2},\frac{1}{2},0\right)^T=0.$$

再计算 $A_2.Y^{\mathrm{T}}$ 在(1,0,0)处的值：

$$(1,-1,-6)(1,0,0)^{\mathrm{T}}=1.$$

比较以上5个值，可知局中人2的最优策略是

$$Y_1^*=(0,1,0),\quad Y_2^*=\left(\frac{1}{2},\frac{1}{2},0\right). \tag{6.22}$$

回到原来的矩阵对策 $\boldsymbol{B}$，根据(6.20)，(6.21)，(6.22)，得到对策 $\boldsymbol{B}$ 的值是 $v_B=3$.

局中人1的全体最优策略是 $\lambda X_1^*+(1-\lambda)X_2^*$，$0\leqslant\lambda\leqslant1$，即

$$\lambda(1,0,0)+(1-\lambda)\left(\frac{1}{3},0,\frac{2}{3}\right),\ 0\leqslant\lambda\leqslant1;$$

局中人2的全体最优策略是 $\mu Y_1^*+(1-\mu)Y_2^*$，$0\leqslant\mu\leqslant1$，即

$$\mu(0,1,0)+(1-\mu)\left(\frac{1}{2},\frac{1}{2},0\right),\ 0\leqslant\mu\leqslant1.$$

### 6.1.9 矩阵对策与线性规划的关系

设 $\boldsymbol{A}=(a_{ij})$ 是 $m\times n$ 矩阵对策的支付矩阵. 可假设对一切 $i$ 和一切 $j$ 有 $a_{ij}>0$，则对策的值 $v>0$；否则只需对 $\boldsymbol{A}$ 的每个元素加上一个适当的正的常数，就可满足上述条件.

当局中人1采用混合策略 $X\in S_m$ 时，他的期望支付至少是

$$\min_{1\leqslant j\leqslant n}XA_{\cdot j}=u.$$

我们得到

$$XA_{\cdot j}\geqslant u,\ j=1,\cdots,n,$$

即

$$\begin{cases}\displaystyle\sum_{i=1}^{m}a_{ij}x_i\geqslant u, & j=1,\cdots,n,\\ \displaystyle\sum_{i=1}^{m}x_i=1, & \\ x_i\geqslant0, & i=1,\cdots,m.\end{cases}$$

令$\frac{x_i}{u}=x'_i, i=1,\cdots,m$，则上面各式化为

$$\begin{cases}\sum_{i=1}^{m}a_{ij}x'_i \geqslant 1, & j=1,\cdots,n,\\ \sum_{i=1}^{m}x'_i=\frac{1}{u}, \\ x'_i \geqslant 0, & i=1,\cdots,m.\end{cases}$$

由于局中人 1 希望使 $u$ 为极大（根据(6.8)，此极大值就是对策的值 $v$），也就是使 $1/u$ 为极小，因此，上述问题化为下列线性规划问题：

$$\begin{aligned}\min \quad & x'_1+\cdots+x'_m;\\ \text{s.t.} \quad & \sum_{i=1}^{m}a_{ij}x'_i \geqslant 1 \quad j=1,\cdots,n,\\ & x'_i \geqslant 0, \quad i=1,\cdots,m.\end{aligned} \tag{6.23}$$

类似地，当局中人 2 采用混合策略 $Y\in S_n$ 时，局中人 1 得到的期望支付不超过

$$\max_{1\leqslant i\leqslant m}\boldsymbol{A}_{i\cdot}\boldsymbol{Y}^{\mathrm{T}}=w.$$

我们得到

$$\boldsymbol{A}_{i\cdot}\boldsymbol{Y}^{\mathrm{T}}\leqslant w, \quad i=1,\cdots,m,$$

即

$$\begin{cases}\sum_{j=1}^{n}a_{ij}y_j \leqslant w, & i=1,\cdots,m,\\ \sum_{j=1}^{n}y_j=1, \\ y_j \geqslant 0, & j=1,\cdots,n.\end{cases}$$

令$\frac{y_j}{w}=y'_j, j=1,\cdots,n$. 由于局中人 2 希望使 $w$ 为极小（这个极小值

就是 $v$)，也就是使 $1/w$ 为极大，因此，上述问题化为下列线性规划问题：

$$\max \quad y'_1 + \cdots + y'_n;$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} y'_j \leqslant 1 \quad i = 1,\cdots,m, \tag{6.24}$$

$$y'_j \geqslant 0, \quad j = 1,\cdots,n.$$

(6.23)和(6.24)是对偶线性规划问题.

由以上分析可知，$m \times n$ 矩阵对策 $\boldsymbol{A}=(a_{ij})$ 的求解问题等价于解上述对偶线性规划问题(6.23)和(6.24).

## 6.2 无限对策

### 6.2.1 零和二人无限对策

矩阵对策最简单的推广，就是将每个局中人的策略集从一个有限集换成一个无限集，例如换成区间[0,1]中的全体实数.

**定义 6.2.1** 局中人 1 从区间[0,1]中选择一个数 $x$，局中人 2 完全独立地从区间[0,1]中选择一个数 $y$. $x$ 和 $y$ 称为局中人 1，2 的纯策略. 选定 $x,y$ 后，局中人 1 得到支付 $P(x,y)$，局中人 2 得到支付 $-P(x,y)$. 这种对策称为零和二人**无限对策**(infinite game). 有时也称它为正方形上的无限对策.

同矩阵对策的情形一样，下面的不等式必定成立：

$$\max_{0\leqslant x\leqslant 1} \min_{0\leqslant y\leqslant 1} P(x,y) \leqslant \min_{0\leqslant y\leqslant 1} \max_{0\leqslant x\leqslant 1} P(x,y), \tag{6.25}$$

假定两端的值都存在.

当且仅当

$$\max_{0\leqslant x\leqslant 1} \min_{0\leqslant y\leqslant 1} P(x,y) = \min_{0\leqslant y\leqslant 1} \max_{0\leqslant x\leqslant 1} P(x,y) \tag{6.26}$$

时，存在点 $(x^*,y^*)\in[0,1]\times[0,1]$，使得不等式

$$P(x,y^*) \leqslant P(x^*,y^*) \leqslant P(x^*,y) \tag{6.27}$$

对于一切 $x\in[0,1]$和一切 $y\in[0,1]$成立.这时,称$(x^*,y^*)$是支付函数 $P(x,y)$或对策的一个(纯策略)鞍点. $P(x,y)$在鞍点处的值 $v=P(x^*,y^*)$称为对策的值.我们有

$$\max_{0\leqslant x\leqslant 1}\min_{0\leqslant y\leqslant 1}P(x,y)=P(x^*,y^*)=v$$

$$=\min_{0\leqslant y\leqslant 1}\max_{0\leqslant x\leqslant 1}P(x,y), \qquad (6.28)$$

$$\max_{0\leqslant x\leqslant 1}P(x,y^*)=P(x^*,y^*)=\min_{0\leqslant y\leqslant 1}P(x^*,y). \qquad (6.29)$$

### 6.2.2 混合策略

如果(6.26)不成立,则(6.25)中的严格不等号成立,即

$$\max_{0\leqslant x\leqslant 1}\min_{0\leqslant y\leqslant 1}P(x,y)<\min_{0\leqslant y\leqslant 1}\max_{0\leqslant x\leqslant 1}P(x,y). \qquad (6.30)$$

这时,就需要引进混合策略的概念.

**定义 6.2.2** 正方形上无限对策局中人 1 的混合策略是定义在[0,1]上的分布函数 $F(x)$: 对于每一个 $x\in[0,1]$,$F(x)$是用某种随机方法选出的数小于或等于 $x$ 的概率,也就是随机变量 $\xi$ 的值小于或等于 $x$ 的概率:

$$F(x)=pr\{\xi\leqslant x\}.$$

当 $x=0$ 时,定义

$$F(0)=pr\{\xi<0\}=0.$$

由定义有

$$F(b)-F(a)=pr\{a<\xi\leqslant b\}, \qquad (6.31)$$

$$F(b)-F(0)=pr\{0\leqslant\xi\leqslant b\}. \qquad (6.32)$$

局中人 2 的混合策略 $G(y)$也是定义在[0,1]上的分布函数.局中人 1,2 分别按分布函数即混合策略 $F(x)$、$G(y)$在区间[0,1]中选择策略.

如果局中人 2 采用纯策略 $y$,局中人 1 采用混合策略 $F(x)$,则局中人 1 的期望支付是

$$\int_0^1 P(x,y)\mathrm{d}F(x),$$

这里的积分是 Stieltjes 积分.

同样,如果局中人 1 采用纯策略 $x$,局中人 2 采用混合策略 $G(y)$,则局中人 1 的期望支付是

$$\int_0^1 P(x,y)\mathrm{d}G(y).$$

如果局中人 1,2 分别采用混合策略 $F(x)$,$G(y)$,则局中人 1 的期望支付是

$$E(F,G) = \int_0^1\int_0^1 P(x,y)\mathrm{d}F(x)\mathrm{d}G(y). \tag{6.33}$$

局中人 1 希望期望支付越大越好,他可以选择 $F(x)$使他得到的期望支付不小于

$$v_1 = \max_F \ \min_G E(F,G), \tag{6.34}$$

这里的最小值和最大值都是在全体分布函数的集上取的.

同样,局中人 2 可以使局中人 1 得到的期望支付不超过

$$v_2 = \min_G \ \max_F E(F,G). \tag{6.35}$$

我们假定 $v_1$,$v_2$ 都存在. 下列不等式必定成立:

$$\begin{aligned} v_1 &= \max_F \ \min_G E(F,G) \\ &\leqslant \min_G \ \max_F E(F,G) = v_2. \end{aligned} \tag{6.36}$$

### 6.2.3 连续对策

在一般情形,(6.36)中的等号不一定成立.

**定理 6.2.3** 设无限对策的支付函数 $P(x,y)$是定义在 $0\leqslant x\leqslant 1$,$0\leqslant y\leqslant 1$ 上的连续函数,则(6.34)和(6.35)存在且相等.

**定义 6.2.4** 称支付函数是连续函数的无限对策为**连续对策**(continuous game).

**定义 6.2.5** 设 $P(x,y)$是正方形 $0\leqslant x\leqslant 1$,$0\leqslant y\leqslant 1$ 上连续

对策的支付函数.如果存在局中人 1,2 的混合策略 $F^*(x)$, $G^*(y)$,使得不等式

$$E(F,G^*) \leqslant E(F^*,G^*) \leqslant E(F^*,G) \tag{6.37}$$

对于一切分布函数 $F$ 和 $G$ 成立,则称$(F^*,G^*)$为 $E(F,G)$或连续对策的一个(混合策略下的)鞍点,或称为对策的一个解. $F^*(x)$, $G^*(y)$分别称为局中人 1,2 的最优(混合)策略.

**定理 6.2.6** 连续对策鞍点存在与定理 6.2.3 等价.

### 6.2.4 最优策略的性质

设 $f(x)$, $g(y)$分别是 $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$ 上的连续函数, $F(x)$, $G(y)$是分布函数,则有

$$\max_F \int_0^1 f(x)\mathrm{d}F(x) = \max_{0\leqslant x\leqslant 1} f(x), \tag{6.38}$$

$$\min_G \int_0^1 g(y)\mathrm{d}G(y) = \min_{0\leqslant y\leqslant 1} g(y). \tag{6.39}$$

利用(6.38)和(6.39)可将关于连续对策的基本定理 6.2.3 写成下列与之等价的形式:

$$\begin{aligned} v_1 &= \max_F \min_{0\leqslant y\leqslant 1} \int_0^1 P(x,y)\,\mathrm{d}F(x) \\ &= \min_G \max_{0\leqslant x\leqslant 1} \int_0^1 P(x,y)\mathrm{d}G(y) = v_2, \end{aligned} \tag{6.40}$$

或

$$\begin{aligned} \max_{0\leqslant x\leqslant 1} \int_0^1 P(x,y)\mathrm{d}G^*(y) &= E(F^*,G^*) \\ &= \min_{0\leqslant y\leqslant 1} \int_0^1 P(x,y)\mathrm{d}F^*(x), \end{aligned} \tag{6.41}$$

其中 $F^*(x)$, $G^*(y)$分别是局中人 1,2 的最优策略.

下面是关于最优策略的两个主要性质,它们与矩阵对策中的定理 6.1.10 和定理 6.1.11 相对应.

**定理 6.2.7** 设 $P(x,y)$是正方形 $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$ 上连续

对策的支付函数；$v$ 是对策的值.

(1) 设 $G^*(y)$是局中人 2 的一个最优策略. 如果对于某个 $x_0 \in [0,1]$有

$$\int_0^1 P(x_0,y)\mathrm{d}G^*(y) < v,$$

则 $pr\{\xi = x_0\} = 0$.

(2) 设 $F^*(x)$是局中人 1 的一个最优策略. 如果对于某个 $y_0 \in [0,1]$有

$$\int_0^1 P(x,y_0)dF^*(x) > v,$$

则 $pr\{\eta = y_0\} = 0$.

**定理 6.2.8** 设 $P(x,y)$是正方形 $0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1$ 上连续对策的支付函数；$v$ 是对策的值. 则

(1) 分布函数 $F^*(x)$是局中人 1 的最优策略的充要条件是

$$\int_0^1 P(x,y)\mathrm{d}F^*(x) \geqslant v,\ 0 \leqslant y \leqslant 1.$$

(2) 分布函数 $G^*(y)$是局中人 2 的最优策略的充要条件是

$$\int_0^1 P(x,y)\mathrm{d}G^*(y) \leqslant v,\ 0 \leqslant x \leqslant 1.$$

### 6.2.5 具凸支付函数的连续对策

如果单位正方形上连续对策的支付函数 $P(x,y)$对于其中一个变量来说是个**凸函数**(convex function)，这种对策称为具凸支付函数的对策. 求解比较简单，解法包含在下述 3 个定理中.

**定理 6.2.9** 设单位正方形上连续对策的支付函数 $P(x,y)$对于每一个 $x$ 是 $y$ 的严格凸函数，则局中人 2 有一个最优纯策略，并且这个纯策略是局中人 2 的唯一最优策略.

**定理 6.2.10** 在定理 6.2.9 的假设条件下，对策的值是

$$v = \min_{0 \leqslant y \leqslant 1} \max_{0 \leqslant x \leqslant 1} P(x,y).$$

由此可知，局中人 2 的最优纯策略 $y^*$ 满足

$$v = \min_{0 \leqslant y \leqslant 1} \max_{0 \leqslant x \leqslant 1} P(x,y) = \max_{0 \leqslant x \leqslant 1} P(x,y^*).$$

**定理 6.2.11** 在定理 6.2.9 的假设条件下，局中人 1 的最优策略是 $F^*(x)$：

(1) 若 $y^*=0$，则

$$F^*(x) = I_{x^*}(x),$$

这里 $I_{x^*}(x)$是具有一个阶梯的阶梯分布函数，即

$$I_{x^*}(x) = \begin{cases} 0, & 0 \leqslant x < x^*, \\ 1, & x^* \leqslant x \leqslant 1, \end{cases}$$

其中 $x^* \in [0,1]$满足条件

$$\begin{cases} P(x^*,0) = v, \\ \dfrac{\partial}{\partial y} P(x^*,0) \geqslant 0. \end{cases}$$

(2) 若 $y^*=1$，则

$$F^*(x) = I_{x^*}(x),$$

其中 $x^* \in [0,1]$满足条件

$$\begin{cases} P(x^*,1) = v, \\ \dfrac{\partial}{\partial y} P(x^*,1) \leqslant 0. \end{cases}$$

(3) 若 $0<y^*<1$，则

$$F^*(x) = \alpha I_{x_1^*}(x) + (1-\alpha) I_{x_2^*}(x), \ 0 \leqslant \alpha \leqslant 1,$$

其中 $x_1^* \in [0,1]$，$x_2^* \in [0,1]$和 $\alpha$ 满足条件

$$\begin{cases} P(x_1^*,y^*) = v, \\ \dfrac{\partial}{\partial y} P(x_1^*,y^*) \geqslant 0, \\ P(x_2^*,y^*) = v, \\ \dfrac{\partial}{\partial y} P(x_2^*,y^*) \leqslant 0, \\ \alpha \dfrac{\partial}{\partial y} P(x_1^*,y^*) + (1-\alpha) \dfrac{\partial}{\partial y} P(x_2^*,y^*) = 0. \end{cases}$$

如果支付函数 $P(x,y)$ 是 $y$ 的凸函数，而不是严格凸函数，则以上 3 个定理仍成立，但这时局中人 2 的最优纯策略通常不再是唯一的了.

### 6.2.6 定时对策

**定义 6.2.12** 定义在单位正方形上的无限对策，如果它的支付函数取下列形式：

$$P(x,y)=\begin{cases}M_1(x,y), & \text{当 } x>y,\\ M_0(x,y), & \text{当 } x=y,\\ M_2(x,y), & \text{当 } x<y,\end{cases}$$

其中 $M_0,M_1,M_2$ 都是连续函数，但 $P(x,y)$ 在正方形的对角线 $x=y$ 上不连续，这种对策称为**定时对策**(games of timing).

现在用一个古典的决斗问题作为例子来说明这种对策. 这是有声决斗的最简单例子.

**例 6.2.13** 假设两个决斗者即局中人 1,2 各有一发子弹. 如果一个局中人开了枪而未命中，则他的对手知道他已用掉了仅有的一发子弹，就可以走到面对面的地方开枪，从而稳操胜券.

假定局中人 1,2 从相隔距离为 1 的地方同时起步迎面走向对方. 每个局中人在决斗开始后任何地点都可以开枪，胜者得到支付为 1，败者为 $-1$，双方同时开枪且都击中对方或都未击中对方时支付为 0.

局中人 1 的策略是选择在双方距离为 $x$ 时开枪，$0\leqslant x\leqslant 1$. 假定**命中率函数**(accuracy function)为 $P_1(x)$，它表示当距离为 $x$ 时击中对方的概率. 局中人 2 的策略是选择一个 $y$ 值，$0\leqslant y\leqslant 1$，在双方相距为 $y$ 时开枪. 以 $P_2(y)$ 表示局中人 2 的命中率. 双方都希望选择一个最合适的时机开枪，过早或过晚都对己方不利；这就是定时对策名称的由来.

上述决斗对策的支付函数是

$$P(x,y)=\begin{cases}1\cdot P_1(x)+(-1)[1-P_1(x)]=2P_1(x)-1, & x>y,\\ 1\cdot P_1(x)[1-P_2(x)]+(-1)[1-P_1(x)]P_2(x) & \\ \qquad =P_1(x)-P_2(x), \quad x=y, & \\ (-1)P_2(y)+1\cdot[1-P_2(y)]=1-2P_2(y), & x<y.\end{cases}$$

在上式中，局中人 1 在双方相距为 $x$ 处开枪. $x>y$ 表示局中人 1 先开枪. $P_1(x)$是局中人 2 被击中的概率，$1-P_1(x)$是局中人 2 未被击中的概率. 因此，$1\cdot P_1(x)$与$(-1)[1-P_1(x)]$之和是局中人 1 在 $x$ 处开枪的期望支付.

$x=y$ 是局中人 1 和 2 同时开枪的情形. 其中，$1\cdot P_1(x)[1-P_2(x)]$表示 2 被击中而 1 未被击中时 1 得到的期望支付；另一项$(-1)[1-P_1(x)]P_2(x)$，则是 2 未被击中而 1 被击中时 1 的期望支付.

最后，$x<y$ 是局中人 2 先开枪的情形.

我们来计算$\max\limits_{0\leqslant x\leqslant 1}\min\limits_{0\leqslant y\leqslant 1}P(x,y)$.

由于 $P_1,P_2$ 都是它们各自的变量的递减函数，所以当 $x<y$ 时有 $1-2P_2(y)>1-2P_2(x)$. 因此

$$\begin{aligned}&\max_{0\leqslant x\leqslant 1}\min_{0\leqslant y\leqslant 1}P(x,y)\\ &=\max_{0\leqslant x\leqslant 1}\min\{2P_1(x)-1,P_1(x)-P_2(x),1-2P_2(x)\}.\end{aligned}$$

令

$$\mu(x)=\min\{2P_1(x)-1,P_1(x)-P_2(x),1-2P_2(x)\},$$

并将 $0\leqslant x\leqslant 1$ 分为 3 个区间如下：

$$A=\{x|P_1(x)+P_2(x)\geqslant 1\},$$
$$B=\{x|P_1(x)+P_2(x)=1\},$$
$$C=\{x|P_1(x)+P_2(x)\leqslant 1\}.$$

则

$$\begin{aligned}&\max_{0\leqslant x\leqslant 1}\min_{0\leqslant y\leqslant 1}P(x,y)=\max_{0\leqslant x\leqslant 1}\mu(x)\\ &=\max\{\max_{x\in A}\mu(x),\max_{x\in B}\mu(x),\max_{x\in C}\mu(x)\}.\end{aligned}$$

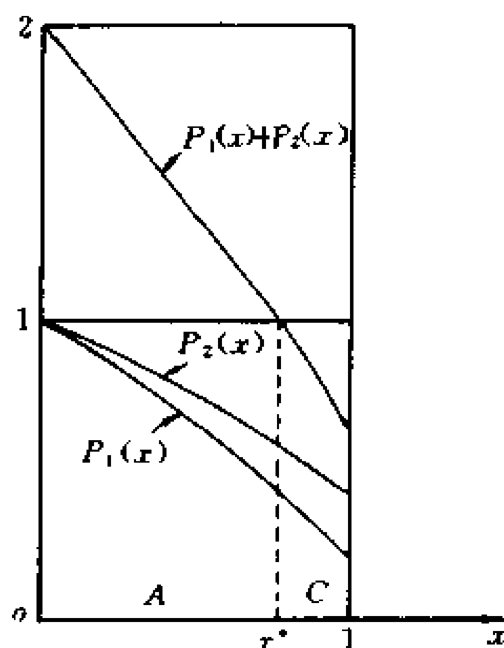



图 6.7

设 $P_1(x^*)+P_2(x^*)=1$，见图 6.7. 则有

$$A=[0,x^*],\ B=\{x^*\},\ C=[x^*,1].$$

(1) 当 $x\in A$ 时，$P_1(x)+P_2(x)\geqslant 1$. 这时有

$$1-2P_2(x)\leqslant P_1(x)-P_2(x)\leqslant 2P_1(x)-1.$$

因此

$$\mu(x)=1-2P_2(x),$$

它是 $x$ 的增函数，见图 6.8. 由此得到

$$\max_{x\in A}\mu(x)=1-2P_2(x^*).$$

(2) 当 $x\in B$ 时，有

$$1-2P_2(x)=P_1(x)-P_2(x)=2P_1(x)-1.$$

因此

$$\mu(x)=P_1(x)-P_2(x),$$

$$\max_{x\in B}\mu(x)=P_1(x^*)-P_2(x^*).$$

(3) 当 $x\in C$ 时，有

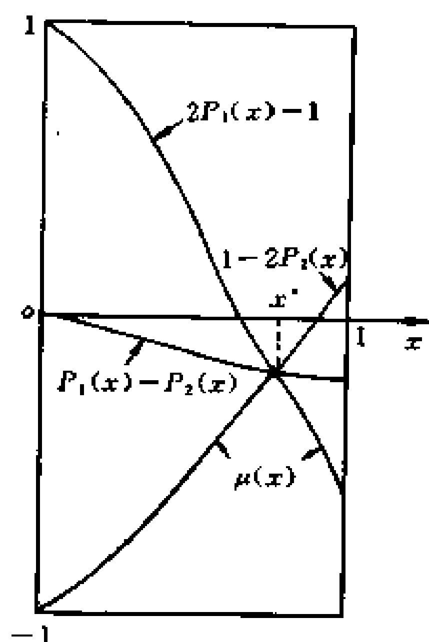



图 6.8

$$2P_1(x)-1 \leqslant P_1(x)-P_2(x) \leqslant 1-2P_2(x).$$

因此

$$\mu(x)=2P_1(x)-1,$$

它是 $x$ 的减函数．由此得到

$$\max_{x\in C}\mu(x)=2P_1(x^*)-1.$$

由上面的(1),(2),(3),有

$$\max_{x\in A}\mu(x)=\max_{x\in B}\mu(x)=\max_{x\in C}\mu(x)$$
$$=P_1(x^*)-P_2(x^*).$$

因此

$$\max_{0\leqslant x\leqslant 1}\min_{0\leqslant y\leqslant 1}P(x,y)=P_1(x^*)-P_2(x^*).$$

类似可证

$$\min_{0\leqslant y\leqslant 1}\max_{0\leqslant x\leqslant 1}P(x,y)=P_1(y^*)-P_2(y^*),$$

其中 $y^*$ 满足

$$P_1(y^*)+P_2(y^*)=1.$$

由此可见，对策有一纯策略鞍点$(x^*,y^*)$，$x^*=y^*$，满足方程

$$P_1(x^*)+P_2(x^*)=1.$$

这就是说，两个决斗者应在相距为$x^*$时同时开枪.对策的值是

$$v=P_1(x^*)-P_2(x^*).$$

## 6.3 n人非合作对策

### 6.3.1 基本概念

**定义 6.3.1** 称

$$\Gamma\equiv[I,\{S_i\},\{P_i\}]$$

为$n$人非合作对策($n$-person noncooperative game). 在上式中$I=\{1,\cdots,n\}$是局中人的集. 每个局中人$i=1,\cdots,n$有一个纯策略的有限集

$$S_i=\{s^{(i)}\}=\{s_1^{(i)},\cdots,s_{m_i}^{(i)}\},$$

其中$s_1^{(i)},\cdots,s_{m_i}^{(i)}$是局中人$i$的$m_i$个纯策略. 每个局中人$i$有一个支付函数$P_i$. 当每个局中人$i$选定一个策略$s^{(i)}$后，就形成了对策的一个纯策略**局势**(situation)

$$\boldsymbol{s}=(s^{(1)},\cdots,s^{(n)}),\quad s^{(i)}\in S_i;$$

支付函数就是局势的函数：

$$P_i=P_i(\boldsymbol{s}),\qquad i=1,\cdots,n.$$

**定义 6.3.2**

$$\boldsymbol{s}\parallel t^{(i)}=(s^{(1)},\cdots,s^{(i-1)},t^{(i)},s^{(i+1)},\cdots,s^{(n)}).$$

上式的意义是：在局势$\boldsymbol{s}=(s^{(1)},\cdots,s^{(n)})$中，第$i$个局中人将他的策略$s^{(i)}$换成$t^{(i)}$，其他局中人的策略不变，这样得到的新的局势就是$\boldsymbol{s}\parallel t^{(i)}$. 显然，$\boldsymbol{s}\parallel s^{(i)}=\boldsymbol{s}$.

**定义 6.3.3** 设$\boldsymbol{s}^*$是$n$人非合作对策$\Gamma\equiv[I,\{S_i\},\{P_i\}]$的一个局势. 如果对于每一个$i\in I$和每一个$s^{(i)}\in S_i$ $(s^{(i)}=s_k^{(i)}$,

$k=1,\cdots,m_i$)，有 $P_i(s^* \| s^{(i)}) \leqslant P_i(s^*)$，则称 $s^*$ 是 $\Gamma$ 的一个**平衡点**(equilibrium point)或平衡局势.

以上 3 个定义都是对纯策略而言的. 显然，在一个 $n$ 人非合作对策中，平衡点不一定存在.

同矩阵对策一样，也需要考虑局中人的混合策略.

对于每一个局中人 $i \in I$，以 $\boldsymbol{x}^{(i)}$ 表示 $i$ 的一个混合策略，即

$$\boldsymbol{x}^{(i)} = (x_1^{(i)}, \cdots, x_{m_i}^{(i)}),$$

其中

$$x_k^{(i)} \geqslant 0,\ k = 1,\cdots,m_i,\quad \sum_{k=1}^{m_i} x_k^{(i)} = 1.$$

局中人 $i$ 以概率 $x_k^{(i)}$ 选择策略 $s_k^{(i)}$，$k=1,\cdots,m_i$.

**定义 6.3.4** 称 $\boldsymbol{x}=(x^{(1)},\cdots,x^{(n)})$ 为对策的一个(混合策略)**局势**，其中 $\boldsymbol{x}^{(i)}=(x_1^{(i)},\cdots,x_{m_i}^{(i)})$ 是局中人 $i$ 的一个混合策略，$i=1,\cdots,n$.

**定义 6.3.5**

$$\boldsymbol{x} \| \boldsymbol{z}^{(i)} = (\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(i-1)},\boldsymbol{z}^{(i)},\boldsymbol{x}^{(i+1)},\cdots,\boldsymbol{x}^{(n)}).$$

上式的意义是：在混合策略局势 $\boldsymbol{x}=(\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(n)})$ 中，局中人 $i$ 将他的混合策略 $\boldsymbol{x}^{(i)}$ 换成另一个混合策略 $\boldsymbol{z}^{(i)}$，其他局中人的策略不变，所得到的局势就是 $\boldsymbol{x} \| \boldsymbol{z}^{(i)}$. 显然，$\boldsymbol{x} \| \boldsymbol{x}^{(i)}=\boldsymbol{x}$.

**定义 6.3.6** 称 $\Gamma \equiv [I,\{X_i\},\{P_i\}]$ 为 $n$ **人非合作对策**(在混合策略意义下)，其中

$$I = \{1,\cdots,n\},$$
$$\{X_i\} = \{X_1,\cdots,X_n\},$$
$$X_i = \{\boldsymbol{x}^{(i)}\} = \{(x_1^{(i)},\cdots,x_{m_i}^{(i)})\},\quad i = 1,\cdots,n,$$
$$x_k^{(i)} \geqslant 0,\ k = 1,\cdots,m_i,\quad \sum_{k=1}^{m_i} x_k^{(i)} = 1;$$
$$\{P_i\} = \{P_1,\cdots,P_n\},$$

$$P_i = P_i(s), \quad i = 1,\cdots,n.$$

为了方便,在两个定义 6.3.1 和 6.3.6 中,用了同一个字母 $\Gamma$ 表示纯策略和混合策略意义下的 $n$ 人非合作对策. 由于混合策略的情况为更一般,这样做不会引起混淆.

**定义 6.3.7** 以 $E_i(x)$表示 $n$ 人非合作对策 $\Gamma=[I,\{X_i\},\{P_i\}]$中局中人 $i$ 在局势 $x$ 下应得的**期望支付**,$i=1,\cdots,n$.

**定义 6.3.8** 设 $x^*$ 是 $n$ 人非合作对策 $\Gamma=[I,\{X_i\},\{P_i\}]$的一个混合策略局势. 如果对于每一个 $i\in I$ 和每一个 $x^{(i)}\in X_i$,有 $E_i(x^*\parallel x^{(i)})\leqslant E_i(x^*)$,则称 $x^*$ 是 $\Gamma$(在混合策略下)的一个**平衡点**或**平衡局势**.

### 6.3.2 平衡点的存在性

**定理 6.3.9** 设 $\Gamma=[I,\{X_i\},\{P_i\}]$是 $n$ 人非合作对策. $x^*$ 是 $\Gamma$ 的平衡点的充要条件是:对于每一个局中人 $i$ 和每一个纯策略 $s^{(i)}\in S_i$,有 $E_i(x^*\parallel s^{(i)})\leqslant E_i(x^*)$. 这里 $E_i(x^*\parallel s^{(i)})$是在 $x^*$ 中将 $i$ 的混合策略 $x^{*(i)}$换成一个纯策略 $s^{(i)}$后局中人 $i$ 的期望支付.

利用这个定理很容易验算某个混合策略局势 $x^*$ 是否对策的平衡点.

**定理 6.3.10** Nash 定理 每一个 $n$ 人非合作对策 $\Gamma=[I,\{X_i\},\{P_i\}]$必有平衡点.

### 6.3.3 2×2 双矩阵对策的平衡点

**定义 6.3.11** 设局中人 1 有 $m$ 个策略 $i=1,\cdots,m$;局中人 2 有 $n$ 个策略 $j=1,\cdots,n$. 设局中人 1 选择策略 $i$,局中人 2 选择策略 $j$ 时,他们得到的支付分别是 $a_{ij}$和 $b_{ij}$,则双方的支付矩阵是

$$\boldsymbol{A}_{m\times n},\boldsymbol{B}_{m\times n}.$$

对策由矩阵 $\boldsymbol{A},\boldsymbol{B}$ 完全确定. 这种对策叫做 $m\times n$ **双矩阵对策**(bimatrix game).

以下只介绍 2×2 双矩阵对策平衡点的求法.

**定理 6.3.12** 设 2×2 双矩阵对策局中人 1,2 的支付矩阵分别是

$$A=\begin{bmatrix} a & b \\ c & d \end{bmatrix},\quad B=\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}.$$

以 $X=(x,1-x)$，$Y=(y,1-y)$分别表示局中人 1,2 的混合策略，其中 $0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1$.

令

$$Q=a-b-c+d,\quad q=d-b, \tag{6.42}$$

$$R=a'-b'-c'+d',\quad r=d'-c'. \tag{6.43}$$

则对策的平衡点根据不同的 $Q,q,R,r$ 值由下面的(1)和(2)两组不等式确定：

(1) 当 $Q=0$ 且 $q=0$ 时，

$$0\leqslant x\leqslant 1,\quad 0\leqslant y\leqslant 1. \tag{6.44}$$

当 $Q=0$, $q>0$ 时，

$$x=0,\quad 0\leqslant y\leqslant 1. \tag{6.45}$$

当 $Q=0$, $q<0$ 时，

$$x=1,\quad 0\leqslant y\leqslant 1. \tag{6.46}$$

当 $Q>0$ 时，

$$\left.\begin{array}{ll} x=0, & y\leqslant q/Q, \\ 0<x<1, & y=q/Q, \\ x=1, & y\geqslant q/Q. \end{array}\right\} \tag{6.47}$$

当 $Q<0$ 时,(6.47)右边第一和第三个关于 $y$ 的不等式换成反方向的不等式.

(2) 当 $R=0$ 且 $r=0$ 时，

$$0\leqslant x\leqslant 1,\quad 0\leqslant y\leqslant 1. \tag{6.48}$$

当 $R=0$, $r>0$ 时，

$$0\leqslant x\leqslant 1,\quad y=0. \tag{6.49}$$

当 $R=0$，$r<0$ 时，

$$0 \leqslant x \leqslant 1, \quad y=1. \tag{6.50}$$

当 $R>0$ 时

$$\left.\begin{array}{ll} x \leqslant r/R, & y=0, \\ x=r/R, & 0<y<1, \\ x \geqslant r/R, & y=1. \end{array}\right\} \tag{6.51}$$

当 $R<0$ 时，(6.51)左边第一和第三个关于 $x$ 的不等式换成反方向的不等式.

将 4 组不等式(6.44)至(6.47)中满足对策条件的一组与(6.48)至(6.51)中满足条件的一组联立起来，即可求得与平衡点相对应的 $x$ 和 $y$ 值.

**例 6.3.13** 囚犯的难题(the prisoners' dilemma)

2×2 双矩阵对策的支付矩阵是

$$A=\begin{bmatrix} 8 & 0 \\ 10 & 2 \end{bmatrix}, \quad B=\begin{bmatrix} 8 & 10 \\ 0 & 2 \end{bmatrix}.$$

利用公式(6.42)和(6.43)计算 $Q,q,R,r$：

$$Q=0,\ q=2>0,$$

$$R=0,\ r=2>0.$$

由(6.45)和(6.49)知，对策的平衡点$(x,y)$满足下列关系：

$$x=0,\ 0 \leqslant y \leqslant 1, \tag{6.52}$$

$$0 \leqslant x \leqslant 1,\ y=0. \tag{6.53}$$

不等式组(6.52)在图 6.9 中以粗黑线条画出，不等式组(6.53)则用虚线画出. 容易看出，粗线和虚线的交点就是对策的平衡点. 因此，对策有唯一的一个平衡点，即

$$x=0, \quad y=0;$$

或

$$(\boldsymbol{X},\boldsymbol{Y})=((0,1),(0,1)).$$

局中人 1 和 2 都采用第 2 个纯策略，得到的支付都是 2.

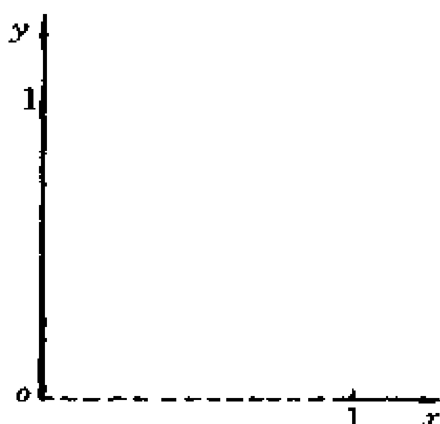



图 6.9

这个结果对两个局中人显然都不是最有利的. 如果双方都选择第 1 个策略，得到的支付都是 8，远较平衡点处的值为大. 但在非合作对策的范畴内，两个局中人没有办法能保证达到这一结局.

**例 6.3.14 周末文娱问题**(battle of the sexes)

2×2 双矩阵对策的支付矩阵是

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

按照公式(6.42)和(6.43)，算得

$$Q = 5 > 0, \quad q = 2,$$
$$R = 5 > 0, \quad r = 3.$$

将这些数值代入公式(6.47)和(6.51)，得到

$$\left.\begin{array}{ll} x = 0, & y \leqslant 2/5, \\ 0 < x < 1, & y = 2/5, \\ x = 1, & y \geqslant 2/5. \end{array}\right\} \qquad (6.54)$$

$$\left.\begin{array}{ll} x \leqslant 3/5, & y = 0, \\ x = 3/5, & 0 < y < 1, \\ x \geqslant 3/5, & y = 1. \end{array}\right\} \qquad (6.55)$$

解这些不等式，求得对策有 3 个平衡点：

$$(x,y)=(0,0),\left(\frac{3}{5},\frac{2}{5}\right),(1,1).$$

用局中人 1,2 的混合策略

$$(\boldsymbol{X},\boldsymbol{Y})=((x,1-x),(y,1-y))$$

表示,就是

$$((0,1),(0,1)),\left(\left(\frac{3}{5},\frac{2}{5}\right),\left(\frac{2}{5},\frac{3}{5}\right)\right),((1,0),(1,0)).$$

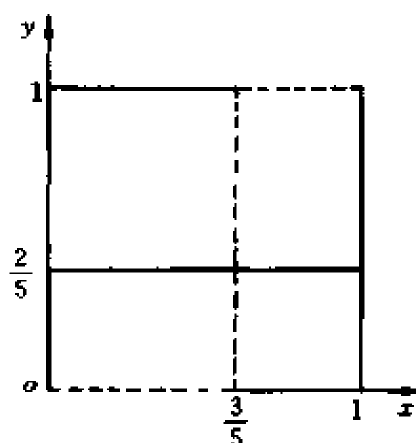



图 6.10

不等式组(6.54)在图 6.10 中以粗黑线条画出,(6.55)则以虚线画出. 容易看出,粗线和虚线的交点就是对策的 3 个平衡点.

如以$(E_1,E_2)$表示局中人 1,2 的期望支付,则在上述 3 个平衡点处的期望支付依次是

$$(E_1,E_2)=(1,2),\quad \left(\frac{1}{5},\frac{1}{5}\right),\quad (2,1).$$

第二个平衡点处的支付显然比第一和第三个平衡点处局中人 1,2 所得的支付都小. 但由于这是一个非合作对策,不许可在事先对如何选择策略进行协商,不许可两个局中人把他们的策略结合起来,所以无法保证达到第一或第三个平衡局势.

在这个例子里,如果局中人 1 采用策略 $\boldsymbol{X}=(x,1-x),0\leqslant x\leqslant 1$;局中人 2 采用策略 $\boldsymbol{Y}=(y,1-y),0\leqslant y\leqslant 1$,他们的期望支付

分别是

$$E_1(x,y) = XAY^{\mathrm{T}} = 5xy - 2(x+y) + 1, \qquad (6.56)$$

$$E_2(x,y) = XBY^{\mathrm{T}} = 5xy - 3(x+y) + 2. \qquad (6.57)$$

现在画出 $E_1E_2$ 平面上当 $0 \leqslant x \leqslant 1$，$0 \leqslant y \leqslant 1$ 时 $(E_1, E_2)$ 的图形.

由(6.56)，(6.57)不难得到(消去 $y$，解出 $x$)

$$0 \leqslant 5(E_1 - E_2)^2 - 2(E_1 + E_2) + 1 \leqslant 5. \qquad (6.58)$$

再利用条件 $0 \leqslant x$ 和 $x \leqslant 1$，还可得到

$$3E_1 - 2E_2 + 1 \geqslant 0, \qquad (6.59)$$

$$2E_1 - 3E_2 - 1 \leqslant 0. \qquad (6.60)$$

先看(6.58). 满足不等式

$$5(E_1 - E_2)^2 - 2(E_1 + E_2) + 1 \geqslant 0$$

的点 $(E_1, E_2)$ 的区域是顶点在(1/4,1/4)，过点(1,2)，(2,1)的抛物线的外侧，即包含原点的一侧.

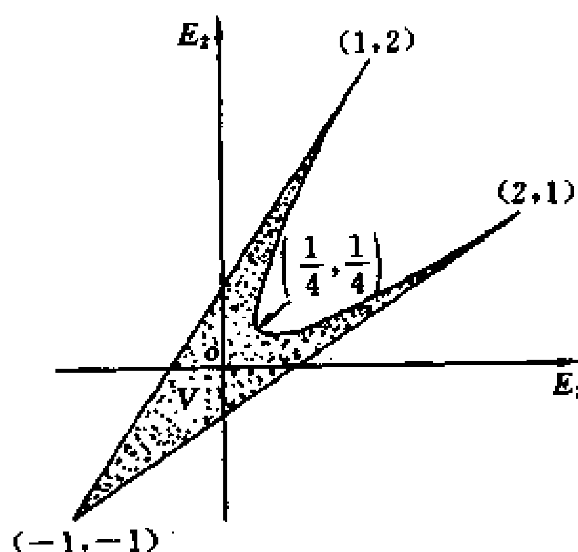



图 6.11

因此，满足(6.58)，(6.59)，(6.60)的点 $(E_1, E_2)$ 形成一个区域，就是图 6.11 中的阴影区域 V. 对于局中人 1,2 的每一对混合

策略$(X,Y)$，有V中一个点$(E_1,E_2)$与之相对应；反之，对于V中每一个点$(E_1,E_2)$，至少有局中人1，2的一对混合策略$(X,Y)$，使得$(E_1,E_2)$是这一对混合策略下局中人1，2的期望支付.通常称区域V为对策的支付集.

## 6.4 $n$人合作对策

### 6.4.1 基本概念、特征函数

在一个$n$人非合作对策中，两个或两个以上的局中人不许可事先商定如何选择策略，不许可把他们的策略结合起来.不允许在局中人之间对得到的支付进行重新分配，一个局中人不能分享另一局中人得到的支付.

$n$人合作对策则对上述两方面的问题都不加限制.局中人可以进行充分的合作：可以事先商定，把他们的策略按任意方式协调结合起来；可以在终局后重新分配若干个局中人所得支付的总和.

因此，一个$n$人合作对策应包含以下几个因素：

首先，若干个局中人要在某些方面进行合作，他们需要结成一个整体，称为联盟或合伙.这是$n$人合作对策的一个重要因素；在非合作对策中，每个局中人都为争取自己最大的利益而奋斗，不存在联合起来进行合作的问题.

其次，对于由局中人结成的每一个联盟，对策应规定一个对应的数量，这个数量可以看作是对联盟的力量或财富或收入大小的一种衡量.这就是说，有一个定义在一切联盟的集上的实值函数.

另外，每个联盟要把得到的总的收入分配给联盟的每一个成员.这就需要用数量来表示这种分配.

在$n$人合作对策中，各个局中人如何选择策略已不是主要问题.强调的乃是联盟的形成.

**定义 6.4.1** 设 $I=\{1,\cdots,n\}$ 是局中人的集，$v(S)$ 是定义在 $I$ 的一切子集，即**联盟**(coalition) $S$ 集上的实值函数，并满足条件

$$v(\varnothing) = 0,$$
$$v(I) \geqslant \sum_{i\in I} v(\{i\}),$$

则称 $\Gamma\equiv[I,v]$ 为 **$n$ 人合作对策**($n$-person cooperative game)，$v(S)$ 为对策的**特征函数**(characteristic function).

在这一节里，假定每个联盟 $S$ 得到的收入可以按照任意方式分配给联盟的成员.这一条件称为**局外支付**(side payments)条件.

**定义 6.4.2** 设 $\Gamma\equiv[I,v]$ 是 $n$ 人合作对策.如果对于一切 $S,T\subset I,S\cap T=\varnothing$，有

$$v(S\cup T) \geqslant v(S)+v(T),$$

则称 $v$ 或 $\Gamma$ 具有**超可加性**(superadditivity).

**定义 6.4.3** 设 $\Gamma\equiv[I,v]$ 是 $n$ 人合作对策.如果对于一切 $S,T\subset I,S\cap T=\varnothing$，有

$$v(S\cup T) = v(S)+v(T),$$

则称 $v$ 或 $\Gamma$ 具有**可加性**(additivity).

**定义 6.4.4** $n$ 人合作对策如果具有可加性，这种对策称为**非实质性的对策**(inessential game)，否则称为**实质性的对策**(essential game).

**定理 6.4.5** $n$ 人合作对策 $\Gamma\equiv[I,v]$ 具有可加性的充要条件是

$$v(I) = \sum_{i\in I} v(\{i\}).$$

主要感兴趣的 $n$ 人合作对策是实质性的对策.由定理 6.4.5 和定义 6.4.1 中特征函数的定义可知，实质性的对策就是特征函数 $v$ 满足条件 $v(I) > \sum_{i\in I} v(\{i\})$ 的对策.

本节中介绍的合作对策，除特别指明者外，假定都是超可

加的.

### 6.4.2 策略等价关系、(0,1)规范化

合作对策可以按其基本性质进行分类,使得每一类对策具有共同的特性.从局中人的角度来考虑,如果两个对策具有共同的局中人,而且他们的特征函数在策略方面的可能性完全相同,就把这两个对策看做是等价的,同时把两个特征函数也看做是等价的.这种等价关系称为策略等价关系.将全体特征函数划分为策略等价类后,从每一类中选出一个最简单的特征函数作为这一类的代表.这样,对整个一类对策或特征函数的研究,就可以通过对这个代表的研究来实现.

**定义 6.4.6** 设 $\Gamma\equiv[I,v]$ 和 $\Gamma'\equiv[I,v']$ 是定义在同一个 $I=\{1,\cdots,n\}$ 上的两个 $n$ 人合作对策.如果存在 $n$ 个常数 $a_1,\cdots,a_n$ 和一个正的常数 $c$,使得对于每一个 $S\subset I$ 有

$$v'(S)=cv(S)+\sum_{i\in S}a_i,$$

则称 $\Gamma$ 和 $\Gamma'$ 是**策略等价**的(strategically equivalent),或者说,特征函数 $v$ 和 $v'$ 是策略等价的.

这样定义的策略等价关系显然满足**等价关系**(equivalence relation)的 3 个条件,即,自反性、对称性和可递性.

**定义 6.4.7** $n$ 人合作对策 $\Gamma\equiv[I,v]$ 的特征函数 $v$ 若满足条件

$$v(\{i\})=0,\quad i=1,\cdots,n$$

和

$$v(I)=1,$$

则称 $\Gamma$ 或 $v$ 是**(0,1)规范化的**(normalized).

**定理 6.4.8** 每一个实质性的 $n$ 人合作对策 $\Gamma\equiv[I,v]$ 策略等价于唯一的一个(0,1)规范化对策.即,对于每一个 $S\subset I$,有

$$v'(S) = cv(S) + \sum_{i \in S} a_i,$$

其中

$$c = \frac{1}{v(I) - \sum_{i \in I} v(\{i\})} > 0,$$

$$a_i = \frac{-v(\{i\})}{v(I) - \sum_{i \in I} v(\{i\})}, \quad i = 1, \cdots, n.$$

$v'(S)$是(0,1)规范化特征函数.

由定理 6.4.8,对于固定的 $I$,合作对策的每一个策略等价类可以由其(0,1)规范化对策作为代表,它具有整个一类对策的共同特性. 因此,对于每一个策略等价类,只需研究其代表即可.

**例 6.4.9** 设 6 人合作对策的特征函数 $v$ 的值是($|S|$是联盟 $S$ 中的局中人数):

$v(S)=1$, 若$|S|=1$,

$v(S)=3$, 若$|S|=2$,

$v(S)=5$, 若$|S|=3$,

$v(S)=7$, 若$|S|=4$,

$v(S)=10$, 若$|S|=5$,

$v(I)=12$.

现在把这个特征函数化为(0,1)规范化形式.

按照定理 6.4.8 中的公式计算 $c$ 和 $a_i$,有

$$c = \frac{1}{v(I) - \sum_{i=1}^{6} v(\{i\})} = \frac{1}{12-6} = \frac{1}{6},$$

$$a_i = -cv(\{i\}) = -\frac{1}{6} \cdot 1 = -\frac{1}{6}, \quad i = 1, \cdots, 6.$$

因此,与 $v$ 策略等价的(0,1)规范化特征函数 $v'$ 的值是

$$v'(\{i\}) = 0, \qquad i = 1, \cdots, 6,$$

$$v'(S) = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}, \quad 若\ |S| = 2,$$

$$v'(S) = \frac{5}{6} - \frac{3}{6} = \frac{2}{6}, \quad 若\ |S| = 3,$$

$$v'(S) = \frac{7}{6} - \frac{4}{6} = \frac{3}{6}, \quad 若\ |S| = 4,$$

$$v'(S) = \frac{10}{6} - \frac{5}{6} = \frac{5}{6}, \quad 若\ |S| = 5,$$

$$v'(I) = 1.$$

对于 $n$ 人合作对策，除了上述(0,1)规范化形式外，有时也用到(−1,0)规范化的形式.

**定义 6.4.10** $n$ 人合作对策 $\Gamma \equiv [I, v]$ 的特征函数 $v$，若满足条件

$$v(\{i\}) = -1, \qquad i = 1, \cdots, n,$$

和

$$v(I) = 0,$$

则称 $\Gamma$ 或 $v$ 是(−1,0)规范化的.

不难验证，从(0,1)规范化特征函数 $v$ 变换成(−1,0)规范化特征函数 $v'$ 的公式是

$$v'(S) = nv(S) - |S|, \qquad S \subset I.$$

在上面的例 6.4.9 中，将(0,1)规范化特征函数 $v'$ 变换成(−1,0)规范化特征函数 $v''$，结果是

$$v''(S) = -1, \quad 若\ |S| \leqslant 4,$$
$$v''(S) = 0, \quad 若\ |S| > 4.$$

**定理 6.4.11** 任何非实质性 $n$ 人合作对策与一个特征函数恒等于零的对策策略等价.

### 6.4.3 二人合作对策

**定义 6.4.12** 在一个 $n$ 人对策中，如果对于每一个局势 $s$ 有

$$\sum_{i=1}^{n} P_i(s) = k,$$

其中 $k$ 为一常数，则称对策为**常和**(constant-sum)对策，否则为**非常和**(non-constantsum)对策.

如果选择 $n$ 个常数 $k_i, i=1,\cdots,n$，使

$$\sum_{i=1}^{n} k_i = k,$$

并令
$$P_i'(s) = P_i(s) - k_i, \quad i = 1,\cdots,n,$$

则
$$\sum_{i=1}^{n} P_i'(s) = 0.$$

可见，常和对策与零和对策从策略上考虑是等价的. 在同一个策略局势下，两者每个局中人得到的支付相差一个常数.

对于特征函数形式下的 $n$ 人合作对策，常和与零和也同样是等价的概念.

首先，关于常和 $n$ 人合作对策的特征函数，有下面的性质.

**定理 6.4.13** 在常和 $n$ 人合作对策 $\Gamma\equiv[I,v]$中，对于每一个联盟 $S\subset I$，有

$$v(S) + v(I\backslash S) = v(I).$$

这一性质称为特征函数的**互补性**(complementarity).

最简单的合作对策是二人合作对策. 如果将二人合作对策分为常和与非常和两类，则根据定理 6.4.13 容易看出，一切常和二人合作对策都是非实质性的.

下面考虑一个非常和二人合作对策.

**例 6.4.14** 在二人非合作对策的例 6.3.14 中，局中人 1,2 的支付矩阵分别是

$$\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}.$$

作为二人非合作对策，已经看到，由于局中人 1,2 只能各自独

立地采用混合策略，所以它的支付集是图 6.11 中的域 V．

现在，如果支付矩阵 $\boldsymbol{A}$，$\boldsymbol{B}$ 所代表的对策是一个合作对策，则局中人 1 和 2 的混合策略可以按任意方式结合起来，形成联合的混合策略．例如，两个局中人可以事先商定，或者两人都选择策略 1，或者都选择策略 2．要做到这一点，只要掷一枚钱币，预先规定，若钱币正面向上，则两人都选择策略 1，若反面向上，则都选择策略 2．

如果允许局中人 1，2 采用一切可能的联合混合策略，则全体支付$(E_1, E_2)$的集是图 6.12 中的区域 V′，它是 3 点(1，2)，(2，1)，(−1，−1)的凸包．

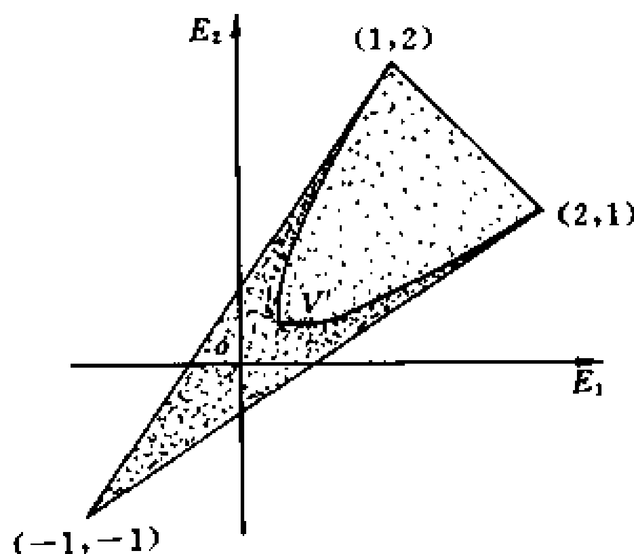



图 6.12

不难算出，这个二人合作对策的特征函数 $v$ 的值是

$$v(\{1\}) = v(\{2\}) = \frac{1}{5}, \quad v(\{1,2\}) = 3.$$

其策略等价于唯一的一个(0，1)规范化对策，其特征函数 $v'$ 的值是

$$v'(\{1\}) = v'(\{2\}) = 0, \; v'(\{1,2\}) = 1.$$

### 6.4.4 转归及其优超关系

**定义 6.4.15** $n$ 维向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$若满足条件

$$x_i \geqslant v(\{i\}), \qquad i = 1,\cdots,n, \tag{6.61}$$

$$\sum_{i=1}^{n} x_i = v(I), \tag{6.62}$$

称为对策 $\Gamma\equiv[I,v]$的一个**转归**(imputation),也可以称为分配.

(6.61)称为**个体合理性**(individual rationality)条件. 可以解释为:一个局中人 $i$,不管他是否参加到一个联盟中去,如果最终分配给他的数额还达不到他一个人单干所能得到的收入,他当然不可能接受这样的分配.

(6.62)称为**集体合理性**(group rationality)或**派雷托最优性**(Pareto optimality)条件. 可以解释为:如果

$$v(I) > \sum_{i=1}^{n} x_i,$$

则 $I$ 中全部成员可以组成大联盟 $I$,得到的总收入 $v(I)$超过转归 $\boldsymbol{x}$ 分配给他们的总数额. 因此,他们肯定不会接受这样的分配方案. 另一方面,$\sum_{i=1}^{n} x_i > v(I)$ 是不可能的,因为总的分配不许可超出总的收入. 由此可见,(6.62)应当成立.

**定理 6.4.16** 非实质性的 $n$ 人合作对策 $\Gamma\equiv[I,v]$只有一个转归,即

$$\boldsymbol{x} = (v(\{1\}),\cdots,v(\{n\})).$$

对于实质性的对策 $\Gamma\equiv[I,v]$,由于

$$a = v(I) - \sum_{i=1}^{n} v(\{i\}) > 0,$$

有无穷多种方式将 $a$ 分为 $n$ 个非负的实数 $a_1,\cdots,a_n$,使得

$$\sum_{i=1}^{n} a_i = a.$$

因此，任何形如

$$x = (v(\{1\}) + a_1, \cdots, v(\{n\}) + a_n)$$

的向量都是 $\Gamma$ 的转归.

对于(0,1)规范化的对策 $\Gamma \equiv [I, v]$，转归 $x = (x_1, \cdots, x_n)$ 应满足的条件(6.61)，(6.62)变为

$$x_i \geqslant 0, \quad i = 1, \cdots, n, \tag{6.63}$$

$$\sum_{i=1}^{n} x_i = 1. \tag{6.64}$$

**定义 6.4.17** $n$ 人合作对策 $\Gamma \equiv [I, v]$ 全体转归的集记为 $X(\Gamma)$.

**定义 6.4.18** 设 $x = (x_1, \cdots, x_n)$ 和 $y = (y_1, \cdots, y_n)$ 是 $n$ 人合作对策 $\Gamma \equiv [I, v]$ 的两个转归，联盟 $S$ 是 $I$ 的非空子集. 如果

$$v(S) \geqslant \sum_{i \in S} y_i, \tag{6.65}$$

且

$$y_i > x_i, \quad i \in S \tag{6.66}$$

则称 $y$ 关于 $S$ **优超**于 $x$($y$ dominates $x$ with respect to $S$)，或者说，$x$ 关于 $S$ 被 $y$ 优超，记为

$$y \succ_S x.$$

(6.65)称为**有效性**(effectiveness)或可行性条件. 或者说，$S$ 是对于 $y$ 的**有效集**(effective set).

一个转归关于某个联盟优超于另一个转归的概念满足可递性. 这就是说，如果 $z \succ_S y$，$y \succ_S x$，则 $z \succ_S x$.

由定义可知，关于单人联盟不可能有转归的优超关系. 关于全体局中人的大联盟 $I$ 也不可能有转归的优超关系.

**定义 6.4.19** 如果存在非空联盟 $S \subset I$，使 $y \succ_S x$，则称转归 $y$ 优超于转归 $x$，记为 $y \succ x$.

转归的一般优超关系不一定满足可递性，举例如下.

设三人合作对策的特征函数 $v$ 的值是

$$v(\{i\}) = 0, \quad i = 1,2,3,$$
$$v(\{1,2\}) = v(\{1,3\}) = v(\{2,3\}) = v(I) = 10.$$
考虑转归
$$z = (0,5,5), \quad y = (6,4,0), \quad x = (4,0,6).$$
显然有 $z \succ y, y \succ x$，但 $z$ 不优超于 $x$.

**定义 6.4.20** 设 $S \subset I, y \in X(\Gamma)$. 定义
$$\text{Dom}_S y = \{x \mid x \in X(\Gamma); y \succ_S x\},$$
称为转归 $y$ 关于联盟 $S$ 的**优超域**(dominion). 并定义
$$\text{Dom } y = \bigcup_{S \subset I} \text{Dom}_S y,$$
称为转归 $y$ 的优超域.

**定义 6.4.21** 设 $S \subset I, A \subset X(\Gamma)$. 定义
$$\text{Dom}_S A = \bigcup_{y \in A} \text{Dom}_S y,$$
称为转归集 $A$ 关于联盟 $S$ 的优超域. 并定义
$$\text{Dom } A = \bigcup_{S \subset I} \text{Dom}_S A,$$
称为转归集 $A$ 的优超域.

以下讨论三人合作对策转归的优超关系.

由于非实质性的合作对策只有一个转归，而且这种对策策略等价于特征函数恒等于零的对策(见定理 6.4.16 和定理 6.4.11)，所以不存在转归的优超问题. 假定所讨论的三人合作对策都是实质性的，且对策已简化为(0,1)规范化形式.

三人合作对策转归集 $X(\Gamma)$的点 $x$ 可用平面上高为 1 的正三角形中重心坐标为$(x_1, x_2, x_3)$的全部点来表示，其中 $x_1, x_2, x_3$ 满足条件
$$x_i \geqslant 0, \quad i = 1,2,3;$$
$$x_1 + x_2 + x_3 = 1.$$
这里的表示方式同图 6.3 一样.

(1) 对于实质性的常和三人合作对策 $\Gamma = [I, v]$，有

$$v(\{i\}) = 0, \qquad i = 1,2,3,$$
$$v(\{1,2\}) = v(\{1,3\}) = v(\{2,3\}) = v(I) = 1.$$
这可从定理 6.4.13 直接推得.

设 $y=(y_1,y_2,y_3)\in X(\Gamma)$. 先考虑联盟{1,2}. 由于 $y \succ_{\{1,2\}} x$ 的定义是

$$v(\{1,2\}) = 1 \geqslant y_1 + y_2, \tag{6.67}$$
$$y_1 > x_1, \qquad y_2 > x_2. \tag{6.68}$$

在图 6.13 中 $y$ 关于联盟{1,2}的优超域是

$$\mathrm{Dom}_{\{1,2\}}y = yb3c, \quad \text{不包括 } yb \text{ 和 } yc.$$

类似地,有

$$\mathrm{Dom}_{\{1,3\}}y = yf2a, \quad \text{不包括 } yf \text{ 和 } ya,$$
$$\mathrm{Dom}_{\{2,3\}}y = yd1e, \quad \text{不包括 } yd \text{ 和 } ye.$$

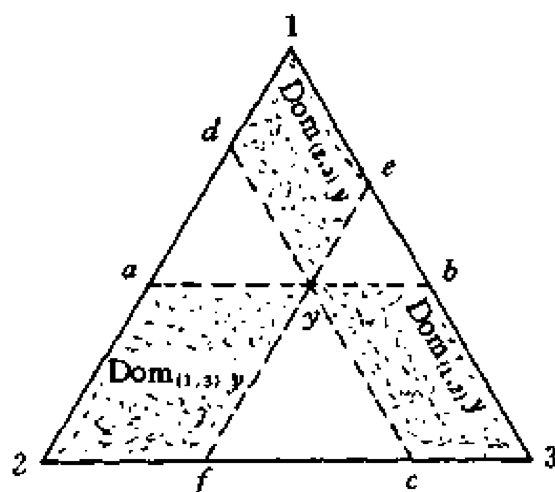



图 6.13

因此,由定义 6.4.20,$y$ 的优超域是

$$\mathrm{Dom}\,y = (\mathrm{Dom}_{\{1,2\}}y) \cup (\mathrm{Dom}_{\{1,3\}}y) \cup (\mathrm{Dom}_{\{2,3\}}y),$$

就是图 6.13 中的三个阴影部分. 图中无阴影部分的三个三角形(包括三角形的边界)是不被转归 $y$ 优超的区域.

(2) 实质性非常和三人合作对策 $\Gamma\equiv[I,v]$特征函数 $v$ 的

值是

$$v(\{i\}) = 0, \qquad i = 1,2,3, \qquad v(I) = 1,$$
$$v(\{1,2\}) = c_3, \quad v(\{1,3\}) = c_2, \quad v(\{2,3\}) = c_1,$$

其中 $c_1, c_2, c_3$ 都是[0,1]中的常数.

$y \succ_{\{1,2\}} x$ 的定义是

$$v(\{1,2\}) = c_3 \geqslant y_1 + y_2, \tag{6.69}$$

$$y_1 > x_1,\ y_2 > x_2. \tag{6.70}$$

有效性条件(6.69)可以改写成

$$y_3 \geqslant 1 - c_3. \tag{6.71}$$

不等式(6.67)是当然成立的.而(6.71)则表示点 $y$ 应位于直线 $y_3 = 1 - c_3$ 上或在其右侧,见图 6.14.

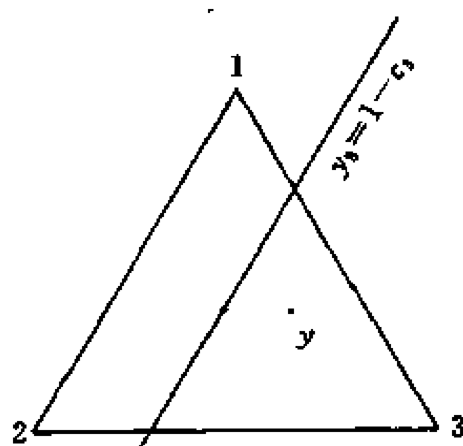



图 6.14

类似地,由 $y \succ_{\{1,3\}} x$ 和 $y \succ_{\{2,3\}} x$ 分别有

$$y_2 \geqslant 1 - c_2, \tag{6.72}$$

$$y_1 \geqslant 1 - c_1. \tag{6.73}$$

一个转归 $y$ 如果满足条件(6.71),(6.72),(6.73),则它的优超域如图 6.15 所示,是三个平行四边形.这与图 6.13 的情况类似.

如果 $y$ 满足条件(6.71)和(6.72),但不满足条件(6.73),则 $y$

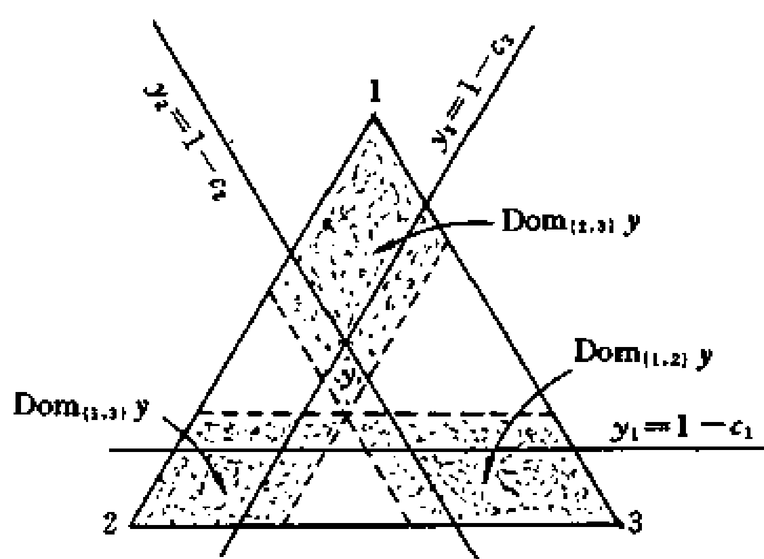



图 6.15

的优超域

$$\mathrm{Dom}\ \boldsymbol{y} = (\mathrm{Dom}_{\{1,2\}}\boldsymbol{y}) \cup (\mathrm{Dom}_{\{1,3\}}\boldsymbol{y}),$$

如图 6.16 中的阴影部分所示.

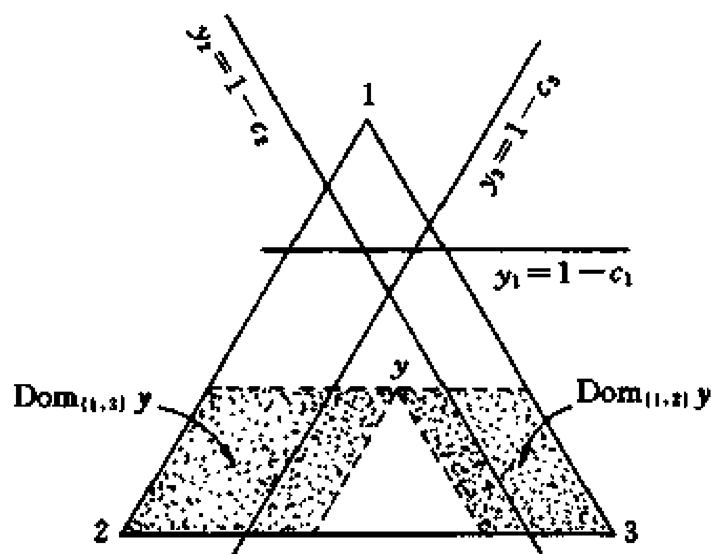



图 6.16

又若 $\boldsymbol{y}$ 满足条件(6.73),但不满足条件(6.71)和(6.72),则

Dom $y$ 是图 6.17 中的一个平行四边形.

其他的情形可以类似地画出,不再一一列举.

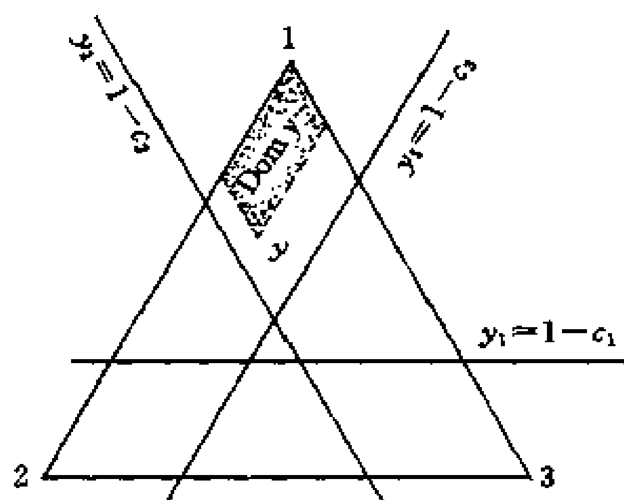



图 6.17

### 6.4.5 核心

为了书写简洁,采用下述记号. 设 $S\neq\varnothing$ 是一个联盟,$\boldsymbol{x}=(x_1,\cdots,x_n)$是一个转归. 记 $x(S)=\sum_{i\in S}x_i$,并规定 $x(\varnothing)=0$.

**定义 6.4.22** 设 $\Gamma=[I,v]$是 $n$ 人合作对策. $\boldsymbol{x}=(x_1,\cdots,x_n)$ 是一个转归,定义

$$C(\Gamma)=\{\boldsymbol{x}\mid\boldsymbol{x}\in X(\Gamma);v(S)-x(S)\leqslant 0,S\subset I\},$$

称为 $\Gamma$ 的**核心**(core).

这个定义表示,对于每一个联盟 $S\subset I$,$C(\Gamma)$中的转归 $\boldsymbol{x}$ 提供给 $S$ 的分配不少于 $S$ 自身所能得到的收入 $v(S)$,因而 $\boldsymbol{x}$ 是能被一切 $S$ 接受的转归.

**定理 6.4.23** 设 $\Gamma=[I,v]$是 $n$ 人合作对策. 转归 $\boldsymbol{x}\in C(\Gamma)$ 的充要条件是 $\boldsymbol{x}$ 不被优超.

我们知道,关于单人联盟$\{i\}$和全体局中人的大联盟 $I$ 都不存在转归的优超关系,所以 $n$ 人合作对策只有当 $n\geqslant 3$ 时才会有

核心.

当 $n\geqslant 3$ 时,任何非实质性的对策 $\Gamma=[I,v]$ 具有可加性,且只有唯一的一个转归(见定义 6.4.4 和定理 6.4.16),即

$$\boldsymbol{x}=(v(\{1\}),\cdots,v(\{n\})),$$

因而这个转归也可以说就是它的核心.

$n\geqslant 3$ 的实质性对策,可以区分为常和 $n$ 人合作对策与非常和 $n$ 人合作对策两类. 关于前者,有下面的定理.

**定理 6.4.24** 设 $\Gamma=[I,v]$ 是实质性的常和 $n$ 人合作对策,则 $C(\Gamma)=\varnothing$.

关于实质性的非常和 $n$ 人合作对策,核心可以是非空的转归集,也可能是空集. 令 $n=3$ 的情形举例如下.

为了简化记号,再规定下列记法:

$$x(12)=x(\{1,2\})=x_1+x_2,$$
$$x(2)=x(\{2\})=x_2,$$
$$v(23)=v(\{2,3\}),\ v(i)=v(\{i\}),$$

等等.

**例 6.4.25** 设三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\quad i=1,2,3,$$
$$v(12)=\frac{2}{3},\ v(13)=\frac{7}{12},\ v(23)=\frac{1}{2},$$
$$v(123)=1.$$

由核心 $C(\Gamma)$ 的定义,转归 $\boldsymbol{x}=(x_1,x_2,x_3)\in C(\Gamma)$ 的充要条件是

$$\begin{cases} v(i)=0\leqslant x_i, \qquad i=1,2,3, \\ v(12)=\dfrac{2}{3}\leqslant x_1+x_2, \\ v(13)=\dfrac{7}{12}\leqslant x_1+x_3, \\ v(23)=\dfrac{1}{2}\leqslant x_2+x_3, \\ v(123)=1=x_1+x_2+x_3. \end{cases}$$

由此得到

$$C(\Gamma)=\left\{\boldsymbol{x}\middle|x_1\leqslant\frac{1}{2},x_2\leqslant\frac{5}{12},x_3\leqslant\frac{1}{3}\right\},$$

它是图 6.18 中的阴影区域，是一个闭三角形.

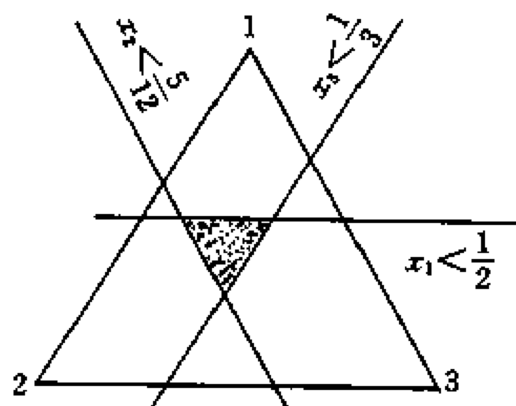



图 6.18

**例 6.4.26** 设三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\quad i=1,2,3,$$

$$v(12)=\frac{1}{3},v(13)=\frac{1}{6},v(23)=\frac{5}{6},$$

$$v(123)=1.$$

$\boldsymbol{x}\in C(\Gamma)$的条件是

$$\begin{cases}v(i)=0\leqslant x_i,\quad i=1,2,3,\\ v(12)=\dfrac{1}{3}\leqslant x_1+x_2,\\ v(13)=\dfrac{1}{6}\leqslant x_1+x_3,\\ v(23)=\dfrac{5}{6}\leqslant x_2+x_3,\\ v(123)=1=x_1+x_2+x_3.\end{cases}$$

因此，

$$C(\Gamma)=\left\{\boldsymbol{x}\middle|x_1\leqslant\frac{1}{6},x_2\leqslant\frac{5}{6},x_3\leqslant\frac{2}{3}\right\},$$

它是一个四边形，见图 6.19.

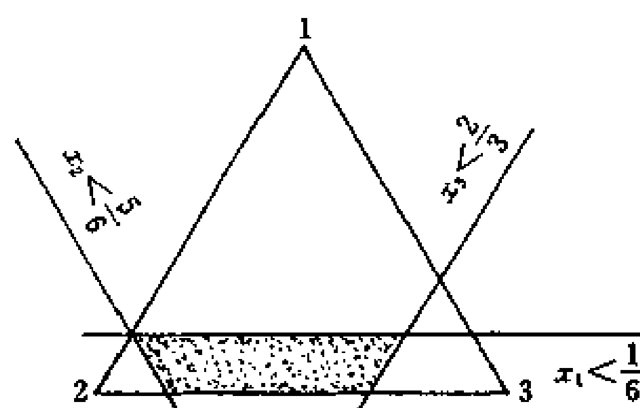



图 6.19

从以上两个例子可以看出，实质性非常和三人合作对策如果具有非空的核心，则这个核心可以是 $X(\Gamma)$ 中的一个点、一个线段、一个三角形、一个四边形、一个五边形或一个六边形.

另一方面，实质性非常和三人合作对策并不一定有非空的核心. 事实上，不难看出，对于(0,1)规范化的非常和三人合作对策来说，核心为空集的充要条件是 $v(12)+v(13)+v(23)>2$.

### 6.4.6 稳定集

上节所介绍的核心无疑是合作对策的一个重要概念. 但如果企图把核心作为合作对策的解，则存在不可克服的困难：有许多对策的核心是空集. 本节介绍一种古典的解的概念.

**定义 6.4.27** 设 $\Gamma=[I,v]$ 是 $n$ 人合作对策；$V\subset X(\Gamma)$ 是满足下面两个条件的转归的集：

(1) 对于任意 $x,y\in V$，有 $x\not\succ y$，

(2) 若 $w\in X(\Gamma)$，$w\notin V$，则存在 $z\in V$ 使 $z\succ w$.

称 $V$ 为对策 $\Gamma$ 的一个**稳定集**(stable set)，或称为 $\Gamma$ 的一个 **vN-M 解**(von Neumann-Morgenstern solution).

定义中的条件(1)表明，$V$ 中的任意两个转归之间没有优超和

被优超的关系．这一性质称为 $V$ 的**内部稳定性**(inner stability)．

条件(2)称为 $V$ 的**外部稳定性**(external stability)．这个性质表明，不在 $V$ 中的每一个转归 $w$，至少被 $V$ 中一个转归 $z$ 所优超．这就是说，至少有一个联盟 $S$ 不喜欢 $w$．这个联盟 $S$ 为了自身的利益希望争取一个分配方案 $z\in V$，使得 $z\succ_S w$．

每一个稳定集是合作对策 $\Gamma$ 在上述意义下的一个解．这种解可以有不止一个，甚至有无穷多个．

合作对策的核心和稳定集之间有下述关系．

**定理 6.4.28** 设 $n$ 人合作对策 $\Gamma\equiv[I,v]$ 有非空的核心 $C(\Gamma)$，且它的 vN-M 解 $V$ 存在，则 $C(\Gamma)\subset V$．

非实质性的合作对策只有一个转归，下面仍然就 $n=3$ 的情形来讨论实质性合作对策的稳定集．同前两段讨论转归的优超关系和核心的程序一样，首先考虑常和三人合作对策，然后再考虑非常和三人合作对策．

(1) 对于实质性的常和三人合作对策 $\Gamma\equiv[I,v]$，有

$$v(i)=0,\quad i=1,2,3,$$
$$v(12)=v(13)=v(23)=v(I)=1.$$

1) 在图 6.20 中，平行于转归三角形一个边的一条直线，例如平行于边 23 的直线段 $bc$，如果它的方程是 $x_1=k, 0\leqslant k<1/2$，则它上面的点所代表的转归具有以下两个性质：

(i) 若 $x, y$ 是 $bc$ 上任意两点，则 $x$ 和 $y$ 互不优超．

(ii) 显然

$$\mathrm{Dom}_{\{2,3\}}bc=bc1,\qquad \text{不包括 } bc,$$
$$\mathrm{Dom}_{\{1,3\}}b=bb'2c,\qquad \text{不包括 } bb' \text{ 和 } bc,$$
$$\mathrm{Dom}_{\{1,2\}}c=cb3c',\qquad \text{不包括 } cc' \text{ 和 } bc.$$

因此

$$\mathrm{Dom}\, bc=X(\Gamma)\backslash bc.$$

由定义可知，$bc$ 是 $\Gamma$ 的一个稳定集．上面的(i)就是稳定集的

内部稳定性,(ii)就是外部稳定性.

当 $k$ 是小于 1/2 的任何非负实数时,这种平行于边 23 的线段 $bc$ 即 $x_1=k$ 有无穷多条,因而 $\Gamma$ 的这一类 vN-M 解有无穷多个.

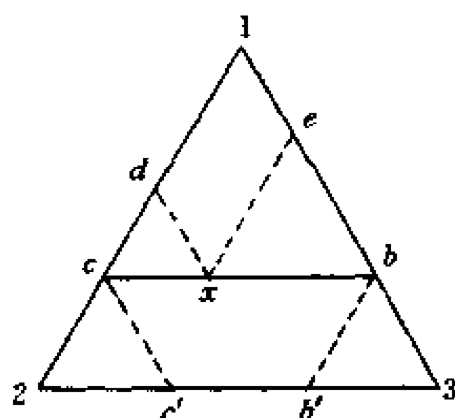



图 6.20

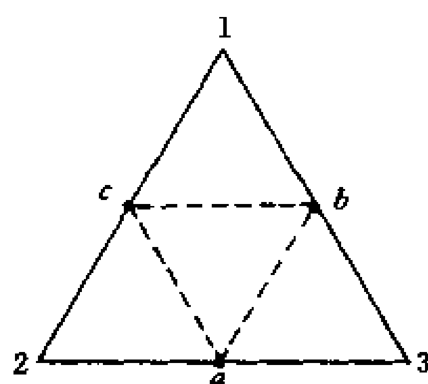



图 6.21

同理,平行于三角形的边 12 或 13 且分别位于另二边中点连线左边或右边的直线段也是 $\Gamma$ 的 vN-M 解.这两类解也都有无穷多个.

2) 常和三人合作对策还有另外一个稳定集,就是转归三角形三条边的三个中点组成的集 $\{a,b,c\}$,见图 6.21.

$V=\{a,b,c\}$ 确为 $\Gamma$ 的一个 vN-M 解,这可以根据定义 6.4.27 来验证:

(i) 内部稳定性. $a,b,c$ 三点中任何二点间没有优超关系.

(ii) 外部稳定性.由于

$$\mathrm{Dom}\ a=\mathrm{Dom}_{\{2,3\}}a=(ac1b)\backslash\{ab,ca\},$$

$$\mathrm{Dom}\ b=\mathrm{Dom}_{\{1,3\}}b=(ba2c)\backslash\{bc,ab\},$$

$$\mathrm{Dom}\ c=\mathrm{Dom}_{\{1,2\}}c=(cb3a)\backslash\{ca,bc\},$$

这就说明了 $X(\Gamma)$ 中除 $a,b,c$ 外任一转归至少被 $a,b,c$ 三者之一优超.

稳定集 $V=\{a,b,c\}=\left\{\left(0,\frac{1}{2},\frac{1}{2}\right),\left(\frac{1}{2},0,\frac{1}{2}\right),\left(\frac{1}{2},\frac{1}{2},0\right)\right\}$

称为 $\Gamma$ 的**对称解**(symmetric solution). 前一种类型即 1)中的解则称为**有差别待遇的解**(discriminatory solution). 得到常数 $k\left(0\leqslant k<\frac{1}{2}\right)$ 的那个局中人称为受到差别待遇的局中人.

(2) 设实质性非常和三人合作对策 $\Gamma=[I,v]$ 的特征函数 $v$ 的值是

$$
\begin{aligned}
&v(i)=0, \quad i=1,2,3,\\
&v(12)=c_3,\ v(13)=c_2,\ v(23)=c_1,\\
&v(I)=1,
\end{aligned}
$$

其中 $0\leqslant c_1,c_2,c_3\leqslant 1$.

以下分为 $C(\Gamma)\neq\varnothing$ 和 $C(\Gamma)=\varnothing$ 两种情形来讨论 $\Gamma$ 的稳定集. 在 6.4.5 核心一节之末曾经指出, $\Gamma$ 的核心为空集的充要条件是

$$
v(12)+v(13)+v(23)>2. \tag{6.74}
$$

1) $C(\Gamma)\neq\varnothing$. 设 $C(\Gamma)$ 如图 6.22 中所示. $C(\Gamma)$ 中的点是不被任何转归优超的转归.

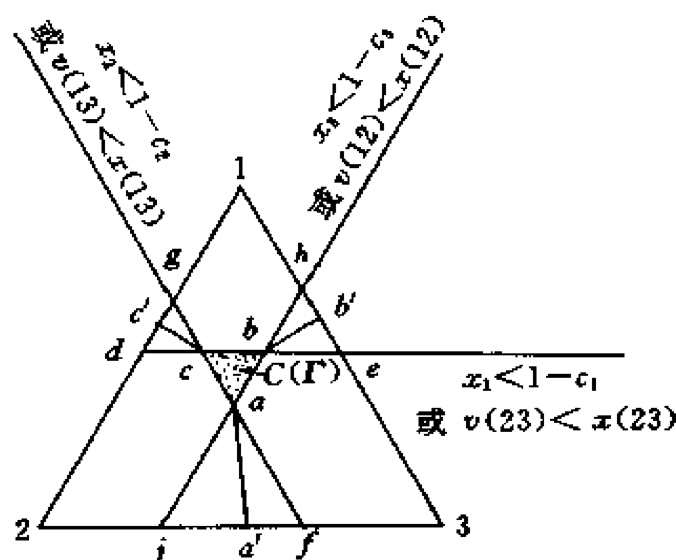



图 6.22

关于联盟{1,2},$C(\Gamma)$中只有线段 $ab$ 上的点满足转归间优超关系的有效性条件(6.65),这时 $v(12)=x(12)$. 显然,

$$\mathrm{Dom}_{\{1,2\}}ab = abe3fa, \qquad \text{不包括 } af, be.$$

同理

$$\mathrm{Dom}_{\{1,3\}}ca = caj2dc, \qquad \text{不包括 } cd, aj.$$

$$\mathrm{Dom}_{\{2,3\}}bc = bcg1hb, \qquad \text{不包括 } bh, cg.$$

由此可见,$C(\Gamma)$中任何两点之间不存在优超关系. $C(\Gamma)$的内部点关于任何二人联盟不满足优超关系的有效性条件. $C(\Gamma)$的边界 $ab$, $bc$, $ca$ 上的点优超于 3 个五边形内部的全部点. 而 3 个三角形 $afj$, $bhe$, $cdg$ 包括其边界的全部点不被 $C(\Gamma)$的边界上的点优超.

试考察 $afj$ 这个三角形. 从顶点 $a$ 到对边 $fj$ 上任一点 $a'$ 作一条连线 $aa'$,则 $aa'$上任意两点互不优超. 这是因为,关于联盟{2,3},$aa'$上的点不满足优超的可行性条件(6.65);而关于联盟{1,2},虽然可行性条件成立,但若 $\boldsymbol{x}=(x_1,x_2,x_3)$, $\boldsymbol{x}'=(x_1',x_2',x_3')$是 $aa'$上的两点,则或者是

$$x_1 > x_1', \qquad x_2 \leqslant x_2',$$

或者是

$$x_1 < x_1', \qquad x_2 \geqslant x_2'.$$

关于联盟{1,3},情况也一样.

再设 $\boldsymbol{x}$ 是 $aa'$上的一点,见图 6.23. 则

$$\mathrm{Dom}_{\{1,2\}}\boldsymbol{x} = \boldsymbol{x}q3s,$$

$$\mathrm{Dom}_{\{1,3\}}\boldsymbol{x} = \boldsymbol{x}r2p.$$

令 $\boldsymbol{x}$ 沿 $aa'$从 $a$ 变到 $a'$,可知 Dom $aa'$包含整个三角形 $afj$ 中除 $aa'$以外的全部点.

同理,从三角形 $C(\Gamma)$的另外两个顶点 $b$, $c$ 分别向小三角形 $bhe$, $cdg$ 的对边上任一点 $b'$, $c'$各作一条连线 $bb'$和 $cc'$,见图 6.22,则 Dom $bb'$包含整个小三角形 $bhe$(除 $bb'$),Dom $cc'$包含整个小三

图　6.23

角形 $cdg$ (除 $cc'$).

综上所述,可得到对策 $\Gamma$ 的一个 vN-M 解:

$$V = C(\Gamma) \cup aa' \cup bb' \cup cc'.$$

这就是说,当非常和三人合作对策 $\Gamma$ 的核心 $C(\Gamma) \neq \varnothing$ 时,根据定理 6.4.28,有 $C(\Gamma) \subset V$. 如果 $C(\Gamma)$ 的图形如图 6.22 中所示,则除 $C(\Gamma)$ 以外再加上三个直线段 $aa'$, $bb'$, $cc'$,就得到 $\Gamma$ 的一个 vN-M 解.

如果 $C(\Gamma)$ 不是一个三角形,而是像图 6.19 中那样的一个四边形,则不被优超的小三角形只有一个,这时只需把核心 $C(\Gamma)$ 加上一条直线段就可以得到 $\Gamma$ 的一个 vN-M 解.

其他情形与此类似,不再列举.

2) $C(\Gamma) = \varnothing$. 前面已经提到,$C(\Gamma) = \varnothing$ 的充要条件是 (6.74). 这时,转归三角形 $X(\Gamma)$ 和直线 $x_1 = 1 - c_1$, $x_2 = 1 - c_2$, $x_3 = 1 - c_3$ 的位置如图 6.24 所示.

小三角形 $abc$ 中没有一个点被它外面的任何点所优超. 并且 $abc$ 中的点关于三个联盟 $\{1,2\}$, $\{1,3\}$, $\{2,3\}$ 都满足优超关系的可行性条件(6.65). 因此,可以像前面(1)中讨论常和三人合作对策那样,考虑小三角形 $abc$ 中平行于 $bc$ 边并且位于 $ab$ 和 $ac$ 中点连线下方的一个线段 $de$,它具有下列性质:

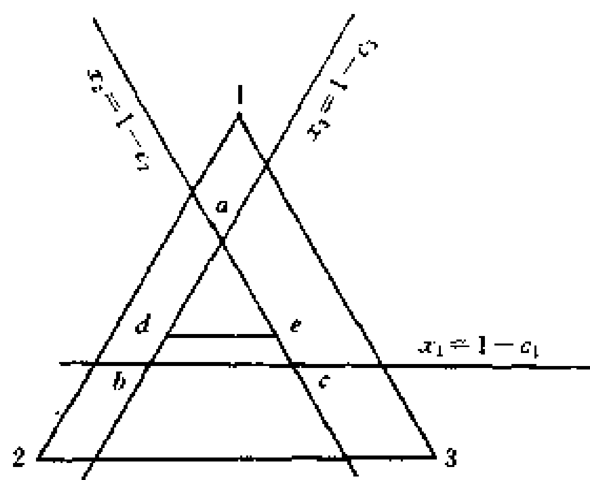



图 6.24

(i) $de$ 上任意两点所代表的转归互不优超.

(ii) $de$ 关于各个二人联盟的优超域是(见图 6.25)

$$\mathrm{Dom}_{\{1,2\}}de=(dh3j)\setminus\{dh,dj\},$$

$$\mathrm{Dom}_{\{1,3\}}de=(em2g)\setminus\{em,eg\},$$

$$\mathrm{Dom}_{\{2,3\}}de=(del1kd)\setminus\{dk,el\}.$$

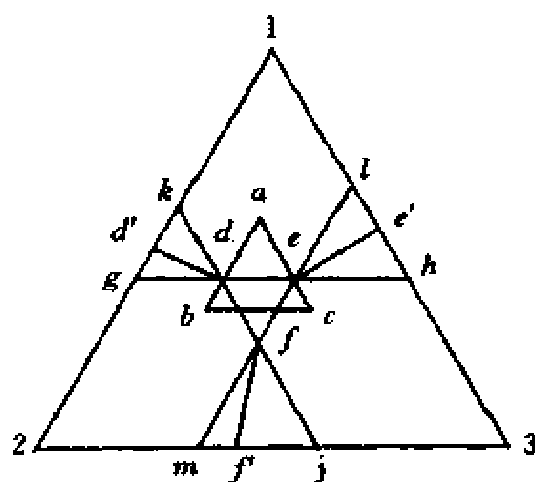



图 6.25

我们得到不被 $de$ 优超的域是三个小三角形 $fjm$, $elh$, $dgk$. 在这三个小三角形中各加上一条直线段 $ff'$, $ee'$, $dd'$, 则 $de$ 和这三个直

线段构成一个 vN-M 解 $V$：

$$V = de \cup ff' \cup ee' \cup dd'.$$

最后，考虑小三角形 $abc$ 三边的三个中点 $d,e,f$. 这三个点互不优超，它们的优超域是（见图 6.26）

$$\mathrm{Dom}_{\{2,3\}}d = (dk1l)\backslash\{dk,dl\},$$

$$\mathrm{Dom}_{\{1,3\}}e = (em2g)\backslash\{em,eg\},$$

$$\mathrm{Dom}_{\{1,2\}}f = (fh3j)\backslash\{fh,fj\}.$$

我们在不被$\{d,e,f\}$优超的三个小三角形 $djm, elh, fgk$ 中各加上一条直线段 $dd', ee', ff'$，则

$$V = dd' \cup ee' \cup ff'$$

也是一个 vN-M 解.

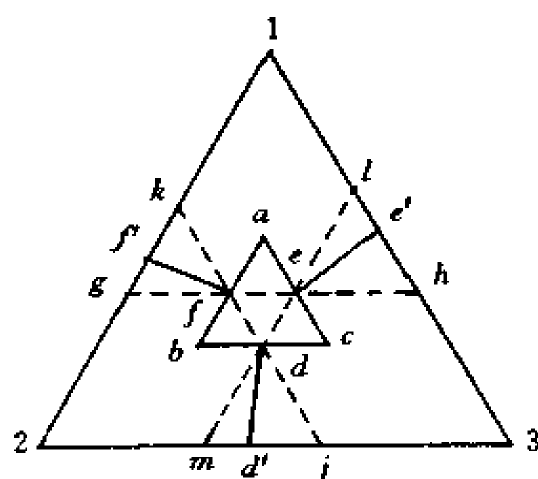



图 6.26

通过以上对三人合作对策的分析，已经足以说明，以稳定集作为合作对策的解的概念，是不能令人满意的.

此外，Lucas 于 1968 年发表了一个 10 人合作对策的例子，它的稳定集不存在；其详细证明发表于 1969 年. 这个例子是

$\Gamma \equiv [I,v], \qquad I = \{1,\cdots,10\},$

$v(I) = 5, \quad v(13579) = 4,$

$v(12) = v(34) = v(56) = v(78) = v(9,10) = 1,$

$$v(3579) = v(1579) = v(1379) = 3,$$
$$v(357) = v(157) = v(137) = 2,$$
$$v(359) = v(159) = v(139) = 2,$$
$$v(1479) = v(3679) = v(5279) = 2,$$
$$v(S) = 0, \qquad \text{对于其他 } S \subset I.$$

这个对策的核心 $C(\Gamma) \neq \varnothing$.

Lucas 和 Rabie 于 1982 年又发表了一个核心为空集而稳定集不存在的 14 人合作对策的例子.

### 6.4.7 广义转归与强 $\varepsilon$ 核心

为了后面两节的需要,这里再引进一些新的概念.

**定义 6.4.29** 设 $\Gamma \equiv [I, v]$是 $n$ 人合作对策. $n$ 维向量 $\boldsymbol{x} = (x_1, \cdots, x_n)$若满足条件 $x(I) = v(I)$,称为一个**广义转归**(pre-imputation). 记全体广义转归的集为 $X^*(\Gamma)$.

$\Gamma$ 的广义转归与定义 6.4.15 中转归的区别在于后者满足个体合理性条件(6.61). $X(\Gamma)$显然是 $X^*(\Gamma)$的子集.

**定义 6.4.30** 设 $\Gamma \equiv [I, v]$是 $n$ 人合作对策. 对于每一个广义转归 $\boldsymbol{x} \in X^*(\Gamma)$和每一个联盟 $S \subset I$,定义 $e(S, \boldsymbol{x}) = v(S) - x(S)$,称为 $S$ 在 $\boldsymbol{x}$ 处的**超出值**(excess).

利用超出值的记号可将核心的定义 6.4.22 改写成

$$C(\Gamma) = \{\boldsymbol{x} \mid \boldsymbol{x} \in X(\Gamma);\ e(S, \boldsymbol{x}) \leqslant 0,\ S \subset I\}. \quad (6.75)$$

这个定义等价于

$$C(\Gamma) = \{\boldsymbol{x} \mid \boldsymbol{x} \in X^*(\Gamma);\ e(S, \boldsymbol{x}) \leqslant 0,\ S \subset I\}. \quad (6.76)$$

**定义 6.4.31** 设 $\Gamma \equiv [I, v]$是 $n$ 人合作对策,$\varepsilon$ 是一个实数. 广义转归的集

$$C_\varepsilon(\Gamma) = \{\boldsymbol{x} \mid \boldsymbol{x} \in X^*(\Gamma); e(S, \boldsymbol{x}) \leqslant \varepsilon, S \subset I, S \neq \varnothing, I\}$$

称为**$\Gamma$ 的强 $\varepsilon$ 核心**(strong ε-core),或简称为 **$\varepsilon$ 核心**.

$\Gamma$ 的核心 $C(\Gamma)$就是强 0 核心 $C_0(\Gamma)$. $\varepsilon$ 当然也可以是负数. 当

$\varepsilon$ 足够小时，$C_\varepsilon(\Gamma)=\varnothing$；当 $\varepsilon$ 充分大时，$C_\varepsilon(\Gamma)\neq\varnothing$. 如果 $\varepsilon_1<\varepsilon_2$，则 $C_{\varepsilon_1}(\Gamma)\subset C_{\varepsilon_2}(\Gamma)$.

**定义 6.4.32** 设 $\Gamma\equiv[I,v]$ 是 $n$ 人合作对策. 若 $\varepsilon_0$ 是使得 $C_\varepsilon(\Gamma)\neq\varnothing$ 的最小 $\varepsilon$，则称 $C_{\varepsilon_0}(\Gamma)$ 为 $\Gamma$ 的**最小核心**(least core)，记为 $LC(\Gamma)$.

$LC(\Gamma)=C_{\varepsilon_0}(\Gamma)$ 是 $\Gamma$ 的一切非空强 $\varepsilon$ 核心之交.

在不致引起混淆的情形，可将 $e(\{2,3\},x)$ 简记为 $e(23)$，$e(\{1\},x)$ 简记为 $e(1)$，等等.

**例 6.4.33** 例 6.4.26 中三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\qquad i=1,2,3,$$

$$v(12)=\frac{1}{3},\quad v(13)=\frac{1}{6},\ v(23)=\frac{5}{6},$$

$$v(123)=1.$$

在图 6.27 中画出了 $\Gamma$ 的核心 $C(\Gamma)=C_0(\Gamma)$ 和强 $\varepsilon$ 核心 $C_\varepsilon(\Gamma)$ 当 $\varepsilon=1/6$ 和 $\varepsilon=2/6$ 的图形. $C(\Gamma)$ 是四边形，$C_{1/6}(\Gamma)$ 是五边形，$C_{2/6}(\Gamma)$ 是六边形. 最小核心是 $C_{-1/12}(\Gamma)$，它是平行于转归三角形 $X(\Gamma)$ 的边 23 的一个直线段，即

$$LC(\Gamma)=C_{-\frac{1}{12}}(\Gamma)=\left\{x\,\middle|\,x_1=\frac{1}{12},x_2\leqslant\frac{9}{12},x_3\leqslant\frac{7}{12}\right\}.$$

**例 6.4.34** 设三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\qquad i=1,2,3,$$

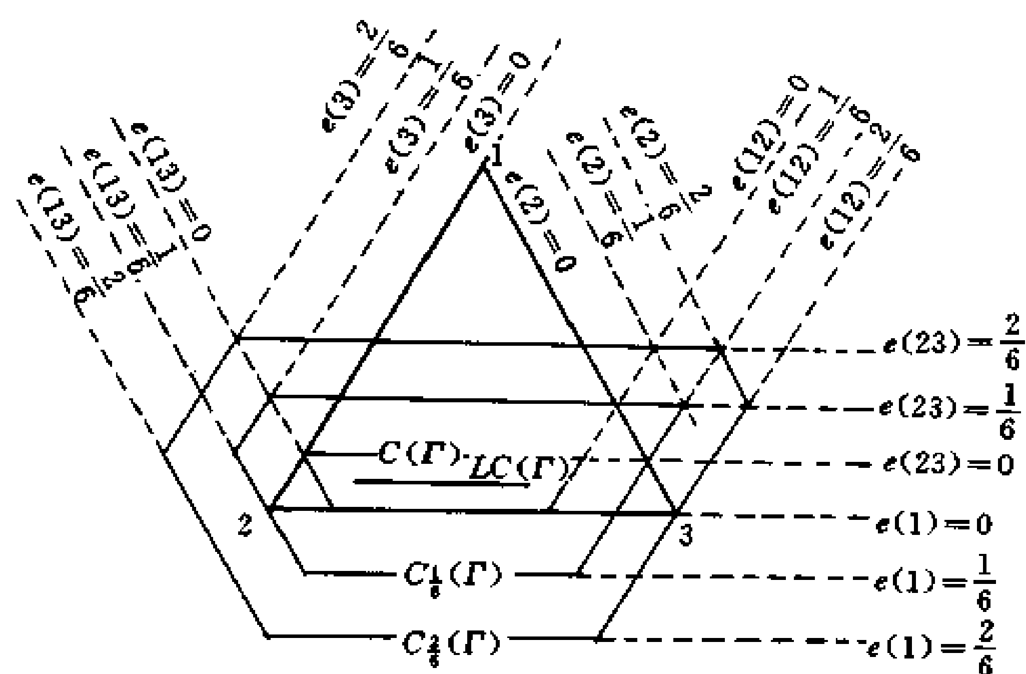
$$v(12)=4,\quad v(13)=3,\quad v(23)=10,$$

$$v(123)=8.$$

特征函数 $v$ 不满足超可加性.

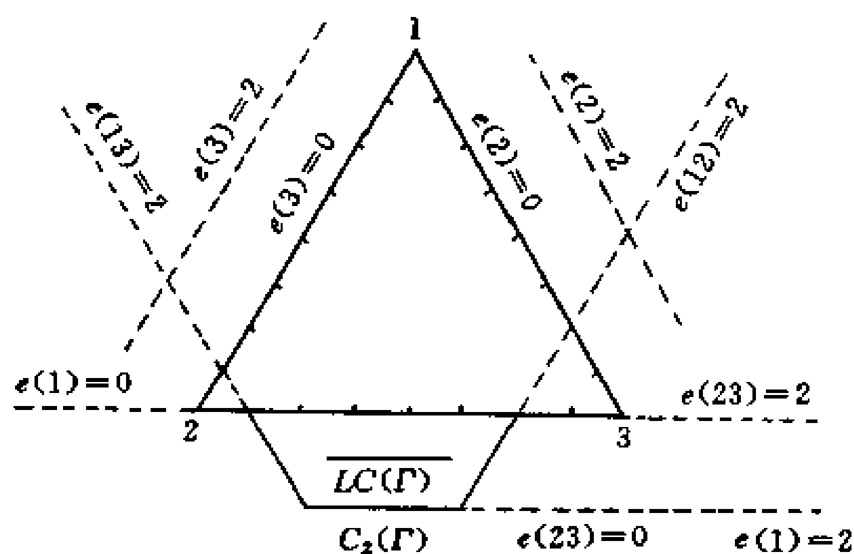
这个对策的核心 $C(\Gamma)$ 是空集.

在图 6.28 中画出了 $\Gamma$ 的强 2 核心 $C_2(\Gamma)$ 的图形，它是由直线 $e(1)=2,e(2)=2,e(3)=2,e(12)=2,e(13)=2,e(23)=2$ 围成的



$e(3)=\frac{2}{6}$
$e(3)=\frac{1}{6}$
$e(3)=0$
$e(13)=0$
$e(13)=\frac{1}{6}$
$e(13)=\frac{2}{6}$
1
$e(2)=0$
$e(2)=\frac{1}{6}$
$e(2)=\frac{2}{6}$
$e(12)=0$
$e(12)=\frac{1}{6}$
$e(12)=\frac{2}{6}$
$e(23)=\frac{2}{6}$
$e(23)=\frac{1}{6}$
$C(\Gamma)$
$\underline{LC(\Gamma)}$
$e(23)=0$
2
3
$e(1)=0$
$C_{\frac{1}{6}}(\Gamma)$
$e(1)=\frac{1}{6}$
$C_{\frac{2}{6}}(\Gamma)$
$e(1)=\frac{2}{6}$


图 6.27



1
$e(13)=2$
$e(3)=2$
$e(3)=0$
$e(2)=0$
$e(2)=2$
$e(12)=2$
$e(1)=0$
$e(23)=2$
2
3
$\overline{LC(\Gamma)}$
$C_2(\Gamma)$
$e(23)=0$
$e(1)=2$


图 6.28

区域，是一个四边形域.

最小核心是

$$LC(\Gamma) = C_1(\Gamma) = \{x \mid x_1 = -1, x_2 \leqslant 6, x_3 \leqslant 5\}.$$

### 6.4.8 核

设 $\Gamma \equiv [I, v]$ 是 $n$ 人合作对策.

**定义 6.4.35** 以 $T_{ij}$ 表示一切包含局中人 $i$ 而不包含局中人 $j$ 的联盟的集，即

$$T_{ij} = \{S \mid S \subset I;\ i \in S,\ j \notin S\}.$$

例如，当 $I = \{1,2,3,4\}$ 时，

$$T_{42} = \{\{4\},\{4,1\},\{4,3\},\{4,1,3\}\}.$$

**定义 6.4.36** 对于每一个广义转归 $x \in X^*(\Gamma)$，定义

$$s_{ij}(x) = \max_{S \in T_{ij}} e(S, x),$$

称为在 $x$ 处局中人 $i$ 超过 $j$ 的**最大剩余**(maximum surplus). 如果

$$s_{ij}(x) > s_{ji}(x),$$

则称在 $x$ 处 $i$ **胜过** $j$($i$ outweighs $j$). 如果在 $x$ 处 $i$ 不胜过 $j$，$j$ 也不胜过 $i$，即，当

$$s_{ij}(x) = s_{ji}(x) \tag{6.77}$$

时，则称 $i$ 和 $j$ 在 $x$ 处**平衡**(in equilibrium).

以上最大剩余、胜过与平衡的概念都是对于广义转归 $x \in X^*(\Gamma)$ 定义的. 对于转归 $x \in X(\Gamma)$，最大剩余的概念完全一样. 胜过与平衡的概念则有所不同.

**定义 6.4.37** 设 $x \in X(\Gamma)$. 如果 $s_{ij}(x) > s_{ji}(x)$，$x_j > v(j)$，则称 $i$ 在 $x$ 处胜过 $j$.

如果 $i$ 不胜过 $j$，$j$ 也不胜过 $i$，则称 $i$ 和 $j$ 在 $x$ 处平衡.

由定义可知，$i$ 和 $j$ 平衡的条件是

$$[s_{ij}(x) - s_{ji}(x)][x_j - v(j)] \leqslant 0, \tag{6.78}$$

$$[s_{ji}(\boldsymbol{x}) - s_{ij}(\boldsymbol{x})][x_i - v(i)] \leqslant 0. \tag{6.79}$$

**定义 6.4.38** $n$ 人合作对策 $\Gamma \equiv [I, v]$ 的**预核**(prekernel)是广义转归 $\boldsymbol{x} \in X^*(\Gamma)$ 的集 $K^*(\Gamma)$，在其中的每一个 $\boldsymbol{x}$ 处，每两个局中人 $i, j$ 关于 $X^*(\Gamma)$ 平衡.

由这个定义可知，广义转归 $\boldsymbol{x} \in K^*(\Gamma)$ 的充要条件是：对于每两个局中人 $i, j$，(6.77)成立.

**定义 6.4.39** $n$ 人合作对策 $\Gamma \equiv [I, v]$ 的**核**(kernel)是转归 $\boldsymbol{x} \in X(\Gamma)$ 的集 $K(\Gamma)$，在其中的每一个 $\boldsymbol{x}$ 处，每两个局中人 $i, j$ 关于 $X(\Gamma)$ 平衡.

由这个定义可知，转归 $\boldsymbol{x} \in K(\Gamma)$ 的充要条件是：对于每一对 $i, j, i \neq j$，不等式(6.78)，(6.79)同时成立. 这等价于

$$s_{ij}(\boldsymbol{x}) = s_{ji}(\boldsymbol{x}), \tag{6.80}$$

或

$$s_{ij}(\boldsymbol{x}) > s_{ji}(\boldsymbol{x}), \quad x_j = v(j), \tag{6.81}$$

或

$$s_{ji}(\boldsymbol{x}) > s_{ij}(\boldsymbol{x}), \quad x_i = v(i). \tag{6.82}$$

比较定义 6.4.38 和定义 6.4.39，预核 $K^*(\Gamma)$ 显然比核 $K(\Gamma)$ 容易计算.

对于较大的 $n$，计算合作对策的核一般是十分复杂的. 以下只通过两个三人合作对策的例子加以说明.

**例 6.4.40** 考虑例 6.4.33 中的三人合作对策 $\Gamma$. 它的特征函数 $v$ 的值是

$$v(i) = 0, \quad i = 1, 2, 3,$$

$$v(12) = \frac{1}{3}, v(13) = \frac{1}{6}, v(23) = \frac{5}{6},$$

$$v(123) = 1.$$

为了计算预核 $K^*(\Gamma)$ 和核 $K(\Gamma)$，对于每一对 $i, j$ 画出方程(6.77)的图形. 例如，为了画出

$$s_{13}(\boldsymbol{x}) = s_{31}(\boldsymbol{x}) \tag{6.83}$$

的图形，先考虑 $s_{13}(\boldsymbol{x})$. 我们知道，

$$s_{13}(\boldsymbol{x}) = \max\{e(1), e(12)\}.$$

在图 6.29 中，以直线 $CD$ 为界，直线的左边 $e(1) > e(12)$，所以 $s_{13}(\boldsymbol{x}) = e(1)$；直线的右边 $e(12) > e(1)$，所以 $s_{13}(\boldsymbol{x}) = e(12)$.

同理，在直线 $AB$ 的左边和右边分别有 $s_{31}(\boldsymbol{x}) = e(3)$ 和 $s_{31}(\boldsymbol{x}) = e(32)$.

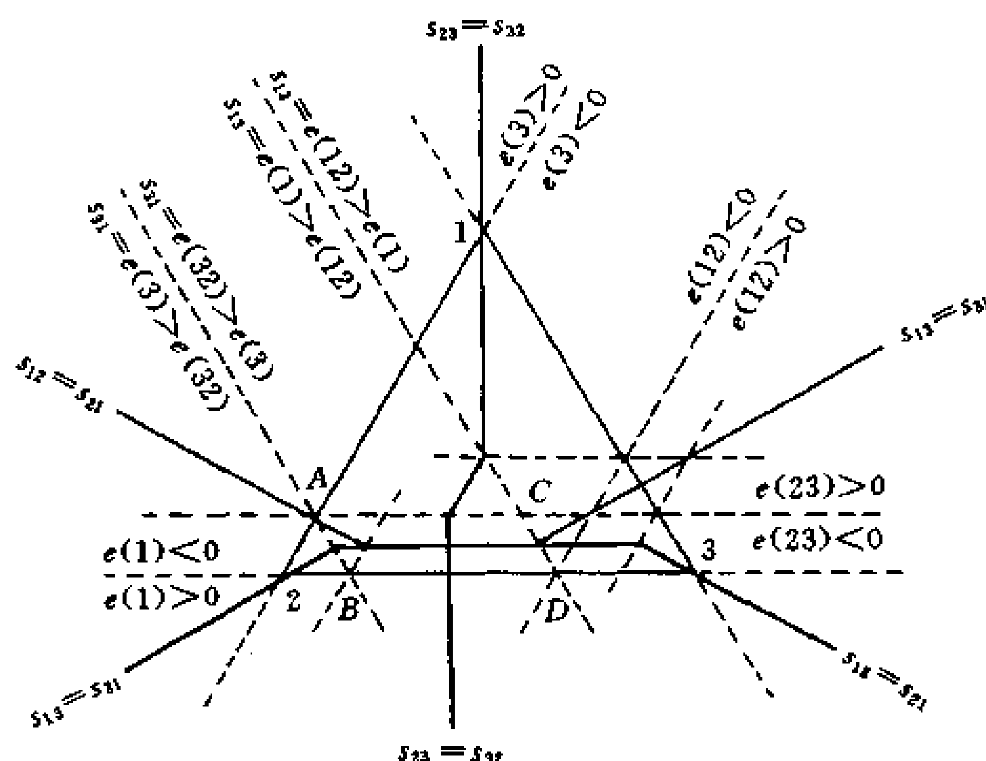



图 6.29

因此，广义转归空间 $X^*(\Gamma)$ 可以分成三个区域. 对于 $AB$ 左边的区域，有

$$s_{13} = e(1), \quad s_{31} = e(3).$$

因而 $s_{13} = s_{31}$ 的图形就是直线 $e(1) = 0$ 和 $e(3) = 0$ 的交角的等分角线. 其次，在直线 $AB$ 和 $CD$ 之间的区域中，

$$s_{13} = e(1), \qquad s_{31} = e(32).$$

因而 $s_{13}=s_{31}$ 的图形是距直线 $e(1)=0$ 和 $e(32)=0$ 等远的直线段. 同理，在 $CD$ 右边的区域中，$s_{13}=s_{31}$ 的图形是 $e(12)=0$ 和 $e(32)=0$ 的交角的等分角线. 这样，就得到了 $s_{13}(x)=s_{31}(x)$ 的整个图形，它是由三条直线段组合起来的折线.

按照同样的方法，可以画出

$$s_{12}(x) = s_{21}(x) \tag{6.84}$$

和

$$s_{23}(x) = s_{32}(x) \tag{6.85}$$

的图形，它们都是一些折线. (6.83), (6.84), (6.85)的图形交于一点. 这就是说，在这一点处，(6.77)或(6.80)对于每一对 $i,j$ 成立. 根据定义 6.4.38 和定义 6.4.39，每两个局中人 $i,j$ 在上述交点处都平衡. 因此，这个交点既属于对策的预核 $K^*(\Gamma)$，又属于它的核 $K(\Gamma)$，并且其他点都不属于二者. 不难算出，

$$K^*(\Gamma) = K(\Gamma) = \left\{x \,\middle|\, x_1 = \frac{1}{12},\ x_2 = \frac{13}{24},\ x_3 = \frac{9}{24}\right\}.$$

这一点是对策的最小核心 $LC(\Gamma)=C_{-\frac{1}{12}}(\Gamma)$ 的中点，也是核心 $C(\Gamma)$ 的几何中心；参看例 6.4.33.

**例 6.4.41** 例 6.4.34 中三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i) = 0, \qquad i = 1,2,3,$$
$$v(12) = 4,\ v(13) = 3,\ v(23) = 10,$$
$$v(123) = 8.$$

前已指出，$C(\Gamma)=\varnothing$. 令 $\varepsilon$ 从 0 逐渐增大，当 $\varepsilon=1$ 时，强 $\varepsilon$ 核心开始出现，它位于转归三角形 $X(\Gamma)$ 的外面，这就是最小核心 $LC(\Gamma)=C_1(\Gamma)$；见图 6.30.

按照同例 6.4.40 中一样的方法画出

$$s_{12}(x) = s_{21}(x), \tag{6.86}$$

$$s_{13}(x) = s_{31}(x), \tag{6.87}$$

$$s_{23}(\boldsymbol{x}) = s_{32}(\boldsymbol{x}) \tag{6.88}$$

的图形.这三个方程的图形交于一点,这一点就是构成对策 $\Gamma$ 的预核 $K^*(\Gamma)$的唯一广义转归.

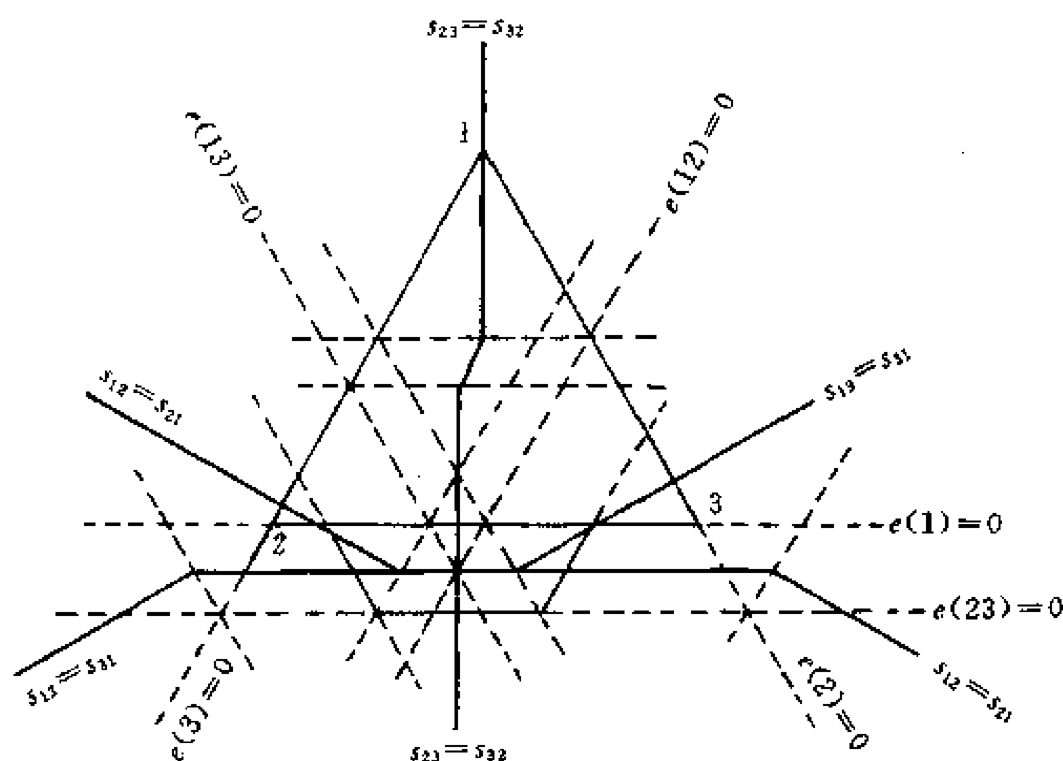



图 6.30

在核 $K(\Gamma)$处,则只有(6.88)一个等式成立.另外两个换成了不等式

$$s_{21}(\boldsymbol{x}) > s_{12}(\boldsymbol{x}), \quad x_1 = 0 = v(1),$$

$$s_{31}(\boldsymbol{x}) > s_{13}(\boldsymbol{x}), \quad x_1 = 0.$$

这两个不等式正是(6.81)或(6.82)的形式.

不难算出,

$$K^*(\Gamma) = \{\boldsymbol{x} \mid x_1 = -1, x_2 = 5, x_3 = 4\}.$$

同例 6.4.40 的情形一样,它也是最小核心 $LC(\Gamma)=C_1(\Gamma)$的中点.

对策的核是

$$K(\Gamma) = \{\boldsymbol{x} | x_1 = 0,\ x_2 = 4.5,\ x_3 = 3.5\}.$$

以上预核和核的概念都是对于全体局中人的集 $I$ 定义的，没有对联盟的构成加上任何限制条件. 以下介绍在一定的联盟结构下对策的核的概念.

**定义 6.4.42** 设 $\mathscr{B}$ 是 $I=\{1,\cdots,n\}$ 的一个**划分**(partition)，即

$$\mathscr{B} = \{B_1,\cdots,B_p\},$$

其中诸 $B_k$ 互不相交，且 $\bigcup_{k=1}^{p} B_k = I$. 称 $\mathscr{B}$ 为**联盟结构**(coalition structure).

**定义 6.4.43** 称

$$(\boldsymbol{x};\mathscr{B}) = (x_1,\cdots,x_n;B_1,\cdots,B_p)$$

为**个体合理支付构形**(individually rational payoff configuration)，其中 $\mathscr{B}$ 是一个联盟结构，$\boldsymbol{x}=(x_1,\cdots,x_n)$ 是一个**支付向量**(payoff vector)，满足条件

$$x_i \geqslant v(i), \qquad i = 1,\cdots,n, \tag{6.89}$$

$$x(B_k) = v(B_k), \qquad k = 1,\cdots,p. \tag{6.90}$$

在一个确定的联盟结构 $\mathscr{B}$ 下，对于满足(6.89)和(6.90)的支付向量 $\boldsymbol{x}$，联盟在 $\boldsymbol{x}$ 处的超出值、最大剩余、胜过及平衡的概念都和前面一样，只是转归向量换成了这里的支付向量，并且在涉及胜过与平衡的概念时，$i$ 和 $j$ 必须属于同一个 $B_k \in \mathscr{B}$. 核的概念则可重新定义如下：

**定义 6.4.44** 设 $\mathscr{B}$ 是 $n$ 人合作对策 $\Gamma=[I,v]$ 的联盟结构. $\Gamma$ 关于 $\mathscr{B}$ 的核 $K(\Gamma)$ 是一切这样的个体合理支付构形 $(\boldsymbol{x};\mathscr{B})$ 的集：每一对属于同一个 $B_k \in \mathscr{B}$ 的局中人 $i,j$ 在 $(\boldsymbol{x};\mathscr{B})$ 处平衡.

如果 $\mathscr{B}=\{I\}$，对策关于全体局中人大联盟的核就是前面定义 6.4.39 中定义的核.

**例 6.4.45** 考虑例 6.4.41 中的三人合作对策 $\Gamma$. 求它关于

联盟结构

$$\{\{2,3\},\{1\}\},\ \{\{1,3\},\{2\}\},\ \{\{1,2\},\{3\}\}$$

的核.

特征函数的值是

$$v(i)=0,\quad i=1,2,3,$$
$$v(12)=4,\ v(13)=3,\ v(23)=10,$$
$$v(123)=8.$$

先看联盟结构 $\mathscr{B}=\{\{2,3\},\{1\}\}$. 由(6.90),有

$$x(1)=x_1=v(1)=0,$$
$$x(23)=x_2+x_3=v(23)=10.$$

可以算出,对策关于联盟结构 $\mathscr{B}=\{\{2,3\},\{1\}\}$的核是

$$\{\boldsymbol{x}|x_1=0,x_2=5.5,x_3=4.5\}.$$

对策关于联盟结构$\{\{1,3\},\{2\}\}$和$\{\{1,2\},\{3\}\}$的核分别是$\{(0,0,3)\}$和$\{(0,4,0)\}$.

前已算出,对策关于全体局中人大联盟的核是

$$\{(0,4.5,3.5)\}.$$

### 6.4.9 核仁

**定义 6.4.46** 设 $\Gamma\equiv[I,v]$是 $n$ 人合作对策,定义 $2^n$ 维向量如下:

$$\theta(\boldsymbol{x})=(\theta_1(\boldsymbol{x}),\cdots,\theta_{2^n}(\boldsymbol{x})),$$

其中

$$\theta_i(\boldsymbol{x})=e(S,\boldsymbol{x}),\quad S\subset I,$$
$$\theta_1(\boldsymbol{x})\geqslant\theta_2(\boldsymbol{x})\geqslant\cdots\geqslant\theta_{2^n}(\boldsymbol{x}).$$

**定义 6.4.47** 设 $\boldsymbol{x},\boldsymbol{y}\in X(\Gamma)$. 若存在下标 $k_0$ 使得

$$\theta_k(\boldsymbol{x})=\theta_k(\boldsymbol{y}),\quad k<k_0,$$
$$\theta_{k_0}(\boldsymbol{x})<\theta_{k_0}(\boldsymbol{y}),$$

则称 $\theta(x)$ 的**字典序**(lexicographic order)在 $\theta(y)$ 之前，或者说，$\theta(x)$ 在字典序上小于 $\theta(y)$，记为 $\theta(x) \prec \theta(y)$.

"非 $\theta(y) \prec \theta(x)$"则记为 $\theta(x) \precsim \theta(y)$.

**定义 6.4.48** $n$ 人合作对策 $\Gamma \equiv [I,v]$ 的**核仁**(nucleolus)是转归 $x$ 的集 $N(\Gamma)$，在其中的每一个 $x$ 处，$\theta$ 在字典序上为最小. 即

$$N(\Gamma) = \{x \mid x \in X(\Gamma);\text{对于一切 } y \in X(\Gamma), \theta(x) \precsim \theta(y)\}. \tag{6.91}$$

核仁的概念是由 Schmeidler 首先提出来的.

**定理 6.4.49** $n$ 人合作对策 $\Gamma \equiv [I,v]$ 的核仁 $N(\Gamma)$ 非空.

**定理 6.4.50** 设 $n$ 人合作对策 $\Gamma \equiv [I,v]$ 的核仁是 $N(\Gamma)$，则 $|N(\Gamma)| = 1$，即，$N(\Gamma)$ 由唯一的一个转归构成.

**定理 6.4.51** 设 $n$ 人合作对策 $\Gamma \equiv [I,v]$ 的核是 $K(\Gamma)$，核仁是 $N(\Gamma)$，则 $N(\Gamma) \subset K(\Gamma)$.

根据这个定理，如果一个 $n$ 人合作对策 $\Gamma$ 的核 $K(\Gamma)$ 只包含一个点，则这个点也就是 $\Gamma$ 的核仁. 例如前面的例 6.4.40 和例 6.4.41 中对策的核都只含有一个点，因而对策的核仁就等于其核.

这个定理不但说明了 $n$ 人合作对策 $\Gamma$ 的核仁 $N(\Gamma)$ 含在核 $K(\Gamma)$ 之中，而且为 $K(\Gamma)$ 的存在性提供了一种证明，比其他直接的证明方法简单得多.

下面是与核仁等价的一个概念.

**定义 6.4.52** 设 $\Gamma \equiv [I,v]$ 是 $n$ 人合作对策. 记

$$X^0 = X(\Gamma),\ \Sigma^0 = \{S \mid S \subset I, S \neq \varnothing, I\}.$$

按照下述方式构造转归集的序列

$$X^0 \supset X^1 \supset \cdots \supset X^\kappa$$

和联盟集的序列

$$\Sigma^0 \supset \Sigma^1 \supset \cdots \supset \Sigma^\kappa.$$

对于 $k = 1, \cdots, \kappa$，定义

$$\epsilon^k = \min_{x \in X^{k-1}} \max_{S \in \Sigma^{k-1}} e(S,x), \tag{6.92}$$

$$X^k = \{x \mid x \in X^{k-1};\ \max_{S \in \Sigma^{k-1}} e(S,x) = \epsilon^k\}, \tag{6.93}$$

$$\Sigma_k = \Big\{S \mid S \in \Sigma^{k-1};\ e(S,x) = \epsilon^k,$$

$$\text{对于一切 } x \in X^k\Big\}, \tag{6.94}$$

$$\Sigma^k = \Sigma^{k-1} \Big\backslash \Sigma_k, \tag{6.95}$$

其中 $\kappa$ 是使得 $\Sigma^k = \varnothing$ 的第一个 $k$ 值.

集 $X^K$ 称为 $\Gamma$ 的**字典中心**(lexicographic center).

**定理 6.4.53** 定义 6.4.52 中的 $\kappa$ 为有限数. 对于 $k=1,\cdots,\kappa$:

(1) $\epsilon^k$ 为有限数;

(2) $X^k$ 为非空紧凸集;

(3) $\Sigma_k \neq \varnothing$;

而对于 $k=1,\cdots,\kappa-1$:

(4) $\epsilon^{k+1}<\epsilon^k$.

**定理 6.4.54** $n$ 人合作对策 $\Gamma\equiv[I,v]$的字典中心由唯一的一个转归构成.

**定理 6.4.55** $n$ 人合作对策 $\Gamma\equiv[I,v]$的核仁等于它的字典中心.

$n$ 人合作对策 $\Gamma\equiv[I,v]$的字典中心的构造为核仁的计算提供了一种方法. 如下所述,可以通过解一系列线性规划来实现.

**算法 6.4.56** 首先,对 $k=1$ 考虑(6.92)～(6.94). 由(6.92)有

$$\epsilon^1 = \min_{x \in X^0} \max_{S \in \Sigma^0} e(S,x).$$

令

$$\max_{S\in\Sigma^0} e(S,\boldsymbol{x}) = \alpha,$$

则

$$\alpha \geqslant e(S,\boldsymbol{x}) = v(S) - x(S), \quad S \in \Sigma^0.$$

即

$$x(S) + \alpha \geqslant v(S), \qquad S \in \Sigma^0.$$

确定 $\varepsilon^1, X^1$ 和 $\Sigma_1$ 的问题等价于解下列线性规划问题：

$$\left.\begin{array}{ll} \min & \alpha; \\ \text{s.t.} & x(S) + \alpha \geqslant v(S), \quad S \in \Sigma^0, \\ & \boldsymbol{x} \in X^0. \end{array}\right\} \tag{6.96}$$

这一线性规划问题的极小值是 $\varepsilon^1$，它是极大超出值的极小值，在一个转归集 $X^1$ 上通过联盟集 $\Sigma_1$ 中的联盟达到这个值 $\varepsilon^1$. 这就是说，对于一切 $S \in \Sigma_1$ 和一切 $\boldsymbol{x} \in X^1$ 有 $e(S,\boldsymbol{x}) = \varepsilon^1$.

其次，我们撇开 $\Sigma_1$ 中的联盟，考虑下列线性规划问题：

$$\left.\begin{array}{ll} \min & \alpha; \\ \text{s.t.} & x(S) + \alpha \geqslant v(S), \quad S \in \Sigma^1 = \Sigma^0 \backslash \Sigma_1, \\ & \boldsymbol{x} \in X^1. \end{array}\right\} \tag{6.97}$$

这一线性规划将给出第二大的超出值 $\varepsilon^2$，以及相应的(6.93)中的 $X^2$ 和(6.94)中的 $\Sigma_2$. 我们再撇开 $\Sigma_2$ 中的联盟，重复以上步骤，直到不再有联盟剩下为止. 最后得到唯一的一个转归，它就是对策 $\Gamma$ 的核仁中的唯一元素.

**例 6.4.57** 我们再一次考虑例 6.4.40 中的三人合作对策 $\Gamma$. 特征函数 $v$ 的值是

$$v(i) = 0, \qquad i = 1,2,3,$$

$$v(12) = \frac{1}{3},\ v(13) = \frac{1}{6},\ v(23) = \frac{5}{6},$$

$$v(123) = 1.$$

第一次线性规划(6.96)给出

$$\alpha = \varepsilon^1 = -\frac{1}{12},$$

$$X^1 = \left\{ \boldsymbol{x} \middle| x_1 = \frac{1}{12},\ x_2 \leqslant \frac{9}{12},\ x_3 \leqslant \frac{7}{12},\ x_2 + x_3 = \frac{11}{12} \right\},$$

$$\Sigma_1 = \{\{1\},\{2,3\}\},$$

$$\Sigma^1 = \Sigma^0 \backslash \Sigma_1 = \{\{2\},\{3\},\{1,2\},\{1,3\}\}.$$

集 $X^1$ 就是 $\Gamma$ 的最小核心.

由第二次线性规划(6.97)得到

$$\alpha = \varepsilon^2 = -\frac{7}{24},$$

$$X^2 = \left\{ \boldsymbol{x} \middle| x_1 = \frac{1}{12},\ x_2 = \frac{13}{24},\ x_3 = \frac{9}{24} \right\},$$

$$\Sigma_2 = \{\{1,2\},\{1,3\}\},$$

$$\Sigma^2 = \Sigma^1 \backslash \Sigma_2 = \{\{2\},\{3\}\}.$$

类似地,由第三次线性规划得到

$$\alpha = \varepsilon^3 = -\frac{9}{24},$$

$$X^3 = X^2,$$

$$\Sigma_3 = \{\{3\}\},$$

$$\Sigma^3 = \Sigma^2 \backslash \Sigma_3 = \{\{2\}\}.$$

而第四次也是最后一次线性规划给出

$$\alpha = \varepsilon^4 = -\frac{13}{24},$$

$$X^4 = X^3 = X^2,$$

$$\Sigma_4 = \{\{2\}\},$$

$$\Sigma^4 = \Sigma^3 \backslash \Sigma_4 = \varnothing.$$

因此,对策的核仁是

$$N(\Gamma) = X^4 = \left\{ \left( \frac{1}{12}, \frac{13}{24}, \frac{9}{24} \right) \right\}.$$

**例 6.4.58** 设 4 人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\qquad i=1,2,3,4,$$
$$v(12)=v(34)=v(14)=v(23)=1,$$
$$v(13)=\frac{1}{2},\qquad v(24)=0,$$
$$v(123)=v(124)=v(134)=v(234)=1,$$
$$v(1234)=2.$$

通过解 3 个线性规划问题就可以求出对策的核仁.容易验证,从

$$X^0=X(\Gamma)\quad 和\quad \Sigma^0=\{S\mid S\subset I;S\neq\varnothing,I\}$$

出发,有

(1) $\varepsilon^1=0$,

$$X^1=\left\{\left(1-\frac{3}{4}\lambda,\frac{3}{4}\lambda,1-\frac{3}{4}\lambda,\frac{3}{4}\lambda\right)\right\},\quad 0\leqslant\lambda\leqslant 1,$$
$$\Sigma_1=\{\{1,2\},\{3,4\},\{1,4\},\{2,3\}\},$$
$$\Sigma^1=\Sigma^0\backslash\Sigma_1$$
$$=\{\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\},\{1,3\},\{2,4\},\{1\},\{2\},\{3\},\{4\}\}.$$

(2) $\varepsilon^2=-\frac{1}{2}$,

$$X^2=\left\{\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)\right\},$$
$$\Sigma_2=\{\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\},\{1,3\},\{1\},\{2\},\{3\},\{4\}\},$$
$$\Sigma^2=\Sigma^1\backslash\Sigma_2=\{\{2,4\}\}.$$

(3) $\varepsilon^3=-1$,

$$X^3=X^2=\left\{\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)\right\},$$
$$\Sigma_3=\{\{2,4\}\},$$
$$\Sigma^3=\Sigma^2\backslash\Sigma_3=\varnothing.$$

因此

$$N(\Gamma) = X^3 = \left\{\left(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)\right\}.$$

### 6.4.10 Shapley 值

Shapley 于 1952 年发现，对于一个 $n$ 人合作对策 $\Gamma$，从 3 条公理出发，可以确定唯一的一组值，这组值可以作为分配给 $\Gamma$ 的全体局中人的值.

**定理 6.4.59** Shapley 值(Shapley value)

设 $\Gamma \equiv [I,v]$是 $n$ 人合作对策. 存在唯一的一组 Shapley 值

$$\Phi_i(v) = \sum_{i\in S}\frac{(n-|S|)!(|S|-1)!}{n!}[v(S)-v(S\backslash i)],$$

$i=1,\cdots,n.$

**例 6.4.60** 例 6.4.57 中三人合作对策 $\Gamma$ 的特征函数 $v$ 的值是

$$v(i)=0,\qquad i=1,2,3,$$

$$v(12)=\frac{1}{3},\quad v(13)=\frac{1}{6},\quad v(23)=\frac{5}{6},$$

$$v(123)=1.$$

这个对策的 Shapley 值是

$$\Phi(v)=\left(\frac{5}{36},\frac{17}{36},\frac{14}{36}\right).$$

**例 6.4.61** 对策论文献中常常引用的一个例子是所谓“投票对策”(voting game). 5 个局中人，局中人 1 有三张投票权，其余的局中人 2,3,4,5 各有一张投票权. 自由结成联盟后，总票数过半即可获胜.

如果把这个合作对策 $\Gamma$ 用(0,1)规范化特征函数 $v$ 表示出来，以 $v=1$ 表示获胜，$v=0$ 表示失败，则

$$v(12)=v(13)=v(14)=v(15)=1,$$

$$v(123)=v(124)=v(125)$$
$$=v(134)=v(135)=v(145)=1,$$
$$v(1234)=v(1235)=v(1245)$$
$$=v(1345)=v(2345)=1,$$
$$v(12345)=1,$$
$$v(S)=0, \quad \text{对于其他的 } S\subset I.$$

容易算出,Shapley 值是

$$\Phi(v)=\left(\frac{6}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10}\right).$$

这个对策关于全体局中人的大联盟的核和核仁是

$$K(\Gamma)=N(\Gamma)=\left\{\left(\frac{3}{7},\frac{1}{7},\frac{1}{7},\frac{1}{7},\frac{1}{7}\right)\right\}.$$

# 参 考 文 献


1 王建华. 对策论,清华大学应用数学丛书第3卷. 清华大学出版社,1986

2 Lucas W F. The Proof that a Game May Not Have a Solution. Trans. Amer. Math. Soc., 1969 137, 219~229

3 Lucas W F and M Rabie. Games with No Solutions and Empty Cores. Math. Op. Res., 1982 7, 491~500

4 Maschler M, B Peleg and L S Shapley. Geometric Properties of the Kernel, Nucleolus, and Related Solution Concepts. Math. Op. Res., 1979 4, 303~338

5 von Neumann J and O Morgenstern. Theory of Games and Economic Behavior. Princeton University Press, Princeton, N. J., 1944,1947,1953 (王建华等译. 竞赛论与经济行为. 科学出版社,1963)

6 Owen G. Game Theory, 2nd ed. New York: Academic Press,1982

7 Schmeidler D. The Nucleolus of a Characteristic Function Game. SIAM J. Appl. Math., 1969, 17, 1163~1170

8 Shapley L S. A Value for n-Person Games. Contributions to the Theory

of Games, Vol. Ⅱ. Annals of Mathematics Studies 28, Kuhn and Tucker, eds. Princeton University Press, Princeton, N. J., 1953, 1956. 307～317

9 Vorob'ev N N. Game Theory, Lectures for Economists and Systems Scientists. New York: Springer-Verlag, 1977

10 Wang Jianhua. The Theory of Games. Oxford Mathematical Monographs. Oxford: Oxford University Press, 1988

# 附录 1　中文—外文名词索引

**E**



**F**

**G**

*H*

**J**

**K**

**Q**



**R**



**S**

**T**

X

Y



Z

# 附录 2　外文—中文名词索引

**A**



**B**

**C**

**D**

**E**

**F**

**J**



**K**



**L**

**M**

**O**



**P**

**T**

# 附录 3　外国人名表

| | | |
|---|---|---|
| A | Abadie, J. | 阿巴迪 |
| B | Balas, E. | 巴拉斯 |
| | Bellman, R. E. | 贝尔曼 |
| | Benayoun, R. | 贝纳扬 |
| | Best, M. J. | 贝斯特 |
| | Broyden, C. G. | 布罗顿 |
| C | Cauchy, A. L. | 柯西 |
| | Charnes, A. | 查尔内斯 |
| | Chvátal, V. | 赫瓦塔尔 |
| D | Daniel, J. | 丹尼尔 |
| | Dantzig, G. B. | 丹齐格 |
| | Davidon, W. C. | 达维顿 |
| | Doig, A. G. | 多伊格 |
| E | Edmonds, J. | 埃德蒙兹 |
| F | Farkas, J. | 法尔卡斯 |
| | Fletcher, R. | 弗莱彻 |
| | Frank, M. | 弗朗克 |
| | Fulkerson, D. R. | 富尔克森 |
| G | Gauss K. F. | 高斯 |
| | Geoffrion, A. M. | 若弗里翁 |
| | Goldfarb, D. | 戈德法布 |

| | | |
|---|---|---|
| | Goldstein, A. A. | 戈尔德施泰因 |
| | Gomory, R. E. | 戈莫里 |
| | Gordan, P. | 戈尔丹 |
| H | Haimes, Y. Y. | 海梅斯 * |
| | Hamilton, W. R. | 汉密尔顿 |
| | Hesse, L. O. | 黑赛 |
| | Hestenes, M. R. | 赫斯特诺斯 * |
| | Hooke, R. | 胡克 |
| | Hu, T. C. | 胡 |
| J | Jacobi, C. G. J. | 雅可比 |
| | Jeeves, T. A. | 吉夫斯 |
| | John, F. | 约翰 * |
| | Johnson, S. M. | 约翰逊松 |
| K | Kelley, J. E. | 凯利 |
| | Kuhn, H. W. | 库恩 |
| L | Lagrange, J. L. | 拉格朗日 |
| | Land, A. H. | 兰德 |
| | Legendre, A. M. | 勒让德尔 |
| | Lemke, C. E. | 莱姆克 |
| | Levenberg, K. A. | 利文贝格 |
| | Little, J. D. C. | 利特尔 |
| | Lucas, W. F. | 卢卡斯 |
| | Marquardt, D. W. | 马夸特 |
| | Morgenstern O. | 莫根斯特恩 |
| N | Nash J. F. | 纳什 |
| | von Neumann, J. | 冯·诺依曼 |

| | | |
|---|---|---|
| | Newton, I. | 牛顿 |
| P | Pareto, V. | 帕雷托 |
| | Polak, E. | 波拉克 |
| | Polyak, B. T. | 波利亚克 |
| | Powell, M. J. D. | 鲍威尔 |
| | Rabie, M. | 拉比耶 |
| | Reeves, C. M. | 里夫斯 |
| | Ribière, G. | 里比埃 |
| | Riccati, J. F. | 里卡蒂 |
| | Rosen, J. B. | 罗森 |
| | Rosenbrock, H. H. | 罗森布罗克 |
| S | Sargent, R. W. H. | 萨金特 |
| | Schmeidler, D. | 施迈德勒 |
| | Schwartz, H. A. | 施瓦茨 |
| | Shannon, D. F. | 香农 |
| | Shapley, L. S. | 沙普利 |
| | Sorenson, H. W. | 索伦森 |
| | Stieltjes, T. J. | 斯蒂尔杰斯 |
| T | Taylor, B. | 泰勒 |
| | Topkis, D. M. | 托普克斯 |
| | Tucker, A. W. | 塔克 |
| V | Veinott, A. F. | 维诺特 |
| W | Wallenius, J. | 瓦勒纽斯 |
| | Wolfe, P. | 沃尔夫 |
| Z | Zionts, S. | 齐翁茨 |

Zoutendijk, G.　周腾狄克

Канторович, Л. В.　康托洛维奇

Марков, А. А.　马尔可夫